序

序的另一层含义是order—有序. 混沌是表象的无序、内在的有序. 什么是混沌? 混沌对不同的人意味着不同的事情. 它是科学的新思维、新前沿.

混沌是对初始值的敏感、是拓扑传递性以及周期点稠密. 混沌是随机性. 混沌是面包师的杰作. 混沌中有奇异吸引子. 混沌的签名是分形. 混沌是周期3. 混沌是正 Lyapunov 指数. 混沌是拓扑熵大于0. 混沌是信息之源, 是信息的膨胀. 混沌是局部的不稳定和整体的稳定. 混沌是简单与复杂的统一. 混沌仅存在于理念之中, 实验中只能得到可以预料的复杂性, 非常像混沌而无法确认其是否真是混沌……

统计学、尤其是非线性时间序列是与动力系统理论及混沌科学有深刻内在联系的. 概率统计是研究随机现象的科学, 而混沌则呈现了随机性, 这使概率统计方法可以作为分析混沌数据的有力工具.

统计学与概率论的方法铺设了走出混沌的路和桥. 深入地了解混沌, 才能走出混沌. 走向数学才能走出混沌.

读了这本小册子之后, 若读者能对上述说法有些体会那就是作者的最大愿望.

方兆本[①]

于中国科技大学

1994.5

① 感谢国家自然科学基金对本书写作的资助.

VISIT…

LANZAROTE Caliente.COM

目　录

前言（王元） 1
序　（方兆本） 3

第一章　偶然与必然 …… 1
§1.1　从掷骰子谈起 …… 1
§1.2　正态分布 …… 9
§1.3　大数定律与中心极限定理 …… 13
§1.4　随机游动与反正弦律 …… 21
第二章　把握随机性 …… 25
§2.1　随机过程和时间序列 …… 26
§2.2　布朗运动 …… 35
§2.3　收敛性与遍历性 …… 42
第三章　混沌 …… 51
§3.1　数学家看混沌 …… 53
§3.2　倍周期、分叉及 Feigenbaum 常数 …… 66
§3.3　奇异吸引子、分形及其维数 …… 71
§3.4　熵—不确定性的度量 …… 89
第四章　走出混沌 …… 96
§4.1　统计与混沌 …… 97
§4.2　分形维数的估计 …… 108
§4.3　非线性时间序列模型 …… 113

§4.4 非参数回归 …………………………………………… 124

附记…………………………………………………………… 128
参考书目……………………………………………………… 129

第一章　偶然与必然

偶然的东西是必然的，必然性本身规定自身为偶然性.

——黑格尔

偶然与必然是人类认识现实世界的一个永恒的话题. 有人以为一件事、一个过程、一种关系不是偶然的便是必然的，非此即彼. 各门学科所研究的当然是带有普遍性的规律. 而能服从普遍规律的事情是必然的. 人们还把他所不能理解、无法解释的现象归因于偶然的或者说超自然的力量：鬼使神差！在本书将要展开的关于混沌的话题里读者将会看到偶然与必然有机统一的有趣模式.

§1.1　从掷骰子谈起

人们对随机现象的研究已有了400多年的历史. 在一门叫

概率统计的学科中研究者们正是力图从偶然中悟出必然．谈概率要从扔骰子说起．机会游戏，赌博在西方叫 Gambling．它要靠扔骰子碰运气．据考古学的发现最早的一颗骰子在伊拉克北部出土，是淡黄色陶器制成的立方体．骰子两对面分别是点子 2 和 3、4 和 5 及 6 和 1．时间约在公元前 2000 多年．虽然早在罗马时代就已有法律禁赌，但数千年来却屡禁不止．在数学著作中讨论赌博的最早例子要算 1494 年威尼斯教士 Luca Pacciolo 的著作中给出点数问题的一个解法．天文学家伽利略曾在 1642 年证明了同时掷 3 个骰子在总共 216 种可能出现的情况中能扔出和为 10 点仅有 27 次．也就是说机会为 27/216=1/8．所谓点数问题是 de Mere 向 Blaise Pascal 提出的．设一个赌徒赌在 8 次扔骰子中掷出一个 6 点．前 3 次他都失败了，赌局因故中断．试问：这时他应分得多少赌注？为此 Pascal 与 Fermat 书信往来研究赌博中的概率问题．什么叫概率，历史上也曾争论不休．一种观点认为它是不确定性的度量；而另一种观点则认为它是人们对某一陈述的信仰程度．以扔骰子出现 6 点为例，概率为 $\frac{1}{6}$．依前者这 $\frac{1}{6}$ 是千百次扔骰子所得经验观察的总结，即所谓统计频率；依后者，人们认为骰子 6 面出现的机会均等，$\frac{1}{6}$ 是对其出 6 点的可能性的认识．科学的发展倒不必等待人们对某一定义的普遍认同，科学的理论甚至有时是建立在定义得并不好的名词之上的．比如，人们对力和加速度的定义不甚满意，但卫星仍依照有关的运动理论在太空飞行．概率也是这样，尽管其定义说不清、道不明，但概率模型和统计方法仍在各个领域找到它自身的位置和价值．本书要讨论的混沌遇到同样的麻烦：什么是混沌？至今没有普遍认可的定义或者说其定义至今还是处于混沌之中！

回到最简单的随机游戏，扔一次骰子就相当于作一次随机试验．试验的结果有 1—6 点 6 种可能，它们组成**样本空间**．样本空间中的点组成**事件**．比如 {2，4，6} 三点组成事件“扔出偶数”．而定义在事件所组成的集合上的衡量事件出现可能性大小的一个集合函数就称为该事件的**概率**．概率作为集合函数其自变量是事件，应变量是 0 和 1 之间的一个实数．它应当满足一定的公理：比如不可能事件的概率为 0，必然发生的事件概率为 1（俗话所说的百分之百的可能）等．本世纪 30 年代，前苏联数学家 Kolmogrov 建立了现代概率论的公理化体系，其数学基础是测度论．测度概念是日常生活中长度、面积、体积概念的自然推广．在本书有关分形维数的讨论中还会碰到测度论的方法与概念．经典概率的计算要依赖于组合计数，通俗而言是要会数数．把总的随机试验的结果数和有利于事件发生的试验结果数数清楚了，它们的比值就是该事件的概率．例如：扔骰子出现偶数的概率为 3/6＝1/2.

科学的惊人成就在于能对未来将要发生的事件进行预报．天文学家可以预报日蚀的准确时间，概率论之所以越来越受到重视也因为它能在大量试验中预报出某一现象出现的频数．当然，在预报之前我们应该先有一个像行星运动规律那样的理论假定．借助于概率的概念来描述的假设称为是统计假设．提出假设、设计实验、收集数据进行分析整理来检验假设、依据检验确认的假设去预测未来，预见未知就是统计学家的工作全貌．统计学家 Karl Pearson 在 1900 年的一篇文章中引用了 Weldon 扔骰子的实例，他把 12 粒骰子扔了 26306 次，对 5 点或 6 点出现的情况作了记录列在表 1.1 中．表中第（1）列为同一次扔的 12 粒中出现 5 或 6 点的粒数，第(4)列为观察频数，第(3)列则为假定骰子是公平的前提下（即各点出现的概率均同为 1/6）所

算出的相应的理论频数或称预报，理论值减去观察值的差列在第(2)列. 首先，人们不难看到理论值与观察值相差不大，相对误差一般只有 0.01—0.10. 且有某种一致性.（3）列理论值的计算要用二项分布. 若作一次试验成功机会为 p 则失败的可能为 $1-p=q$，独立地（这又是统计学中颇费口舌才能解释清楚的概念）作同样的试验就构成一个 Bernoulli 试验序列. 在 n 次 Bernoulli 试验中成功 k 次概率为 $\binom{n}{k}p^kq^{n-k}$，成功次数所服从的是二项分布. 因 $(p+q)^n$ 的二项式展开第 k 项的值与 k 次成功的概率相同而得名. 现在出现 5 点或 6 点的概率为 2/6=1/3，即 $p=\frac{1}{3}$. 于是 12 粒骰子有 2 粒为 5 点或 6 点的概率为

$$P\{\text{成功 2 次}\}=\binom{12}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{10}=\frac{67584}{531441}.$$

(1) 5 或 6 的数目	(2) 超过 I	(3) 预报 I	(4) 观察	(5) 预报 II	(6) 超过 II
0	+18	203	185	187	+2
1	+67	1216	1149	1146	−3
2	+80	3345	3265	3215	−50
3	+101	5576	5475	5465	−10
4	+159	6273	6114	6269	+115
5	−176	5018	5194	5115	−79
6	−140	2927	3067	3043	−24
7	−76	1255	1331	1330	−1
8	−11	392	403	424	+21
9	−18	87	105	96	−9
10	−1	13	14	15	+1
11	−3	1	4	1	−3
12	0	0	0	0	0
总计	0	26，306	26，306	26，306	0

表 1.1

而在 26306 次试验中的近似理论频数应为 26306×P {12 粒中有 2 粒为 5 或 6 点}=3345 次.

细心的读者还会发现在第(2)列(3)—(4)的误差值中呈现出系统的偏差，上面为+下面为−．一般而言，+、−误差应无规律地掺合呈随机的（又一个不易说清的概念）状态．一串+、−号是否随机的检验将在 § 4.1 讨论．但是现在随机性无处可寻、原因何在？人们便怀疑骰子公平的假设是否真实．依据 12×26306 次掷骰子的实际观察人们可以估出出现 5 点或 6 点的概率 $\hat{p}=0.3376986$，$\hat{p}>\frac{1}{3}$．依 $\hat{p}$ 可以算出新的理论频数列第(5)列，再与观察频数(4)列的数据比较，(5)—(4)列入(6)列．这时，可以发现不仅误差减小了，而且+、−相间呈无规律状，随机性更强了．正是这样人们通过建立近似的理论模型来更好地描述客观现实，从偶然中探索必然规律．

读者容易找到各种程度的概率论入门书去了解概率的基本概念，这里引入一些最必须的记号和概念．

定义　样本空间 Ω 的子集合 A（即事件）上的函数 P 若满足 (i) $0\leqslant P(A)\leqslant 1$；

(ii) $P(\varnothing)=0$，$P(\Omega)=1$($\varnothing$ 为空集)；

(iii) 若 A_1，A_2，…为一串互不相交的子集，则

$$P\Big(\bigcup_{k=1}^{\infty}A_k\Big)=\sum_{k=1}^{\infty}P(A_k)$$

称之为**概率**（或概率测度）．

iii)中$\cup$表示集合的并，代表两件事至少发生一件，A_i 互不相交则表示事件不同时发生．iii)又常称作概率的可加性．

定义　随机试验结果的全体称为**样本空间 Ω**，样本空间中

的集合称为**事件**，可以定义概率的事件的全体集合记为$\mathscr{B}$，(Ω，$\mathscr{B}$，P) 称为**概率空间**（它是一般测度空间的特例）.

(Ω，$\mathscr{B}$，P) 便是我们讨论问题的出发点. 为了能比较方便地讨论概率，我们引入随机变量的概念. 随机变量是指其取值要随机会而变的那种变量，如果 $\omega\in\Omega$ 为试验结果则取值为实数的变量 $X(\omega)$ 为随机变量. 而 $X(\omega)$ 在某一范围取值则代表了 $\mathscr{B}$ 中的一个事件. 例如，记

$$X(\omega)=\begin{cases}1, & \text{扔币出正面；}\\ 0, & \text{扔币出反面，}\end{cases}$$

则 $\left\{X(\omega)\leqslant\frac{1}{2}\right\}$、$\{X(\omega)=0\}$，$\{0\leqslant X(\omega)<0.9\}$ 都代表了同一事件{反面向上}. 引入随机变量之后对偶然事件发生的可能性的描述转化为对随机变量的分布的研究. 出现正、反面机会均等就是 $X(\omega)$ 服从两点分布，其 $P\{X=1\}=p$ 和 $P\{X=0\}=1-p=q$，而且 $p=q=\frac{1}{2}$. 研究随机变量既要考查它取值的范围，又要了解取这些值的可能性的大小. 这两点的综合就描述为随机变量的分布. 分布本身即代表偶然现象的内在规律. 随机变量的分布常分为离散的与连续的两大类，后者又有绝对连续与奇异之分. 具体运算中更为方便的是分布函数.

定义 对实数 x，概率 $P\{X(\omega)\leqslant x\}$ 称为是随机变量 X 的**累积分布函数**. 记为 $F(x)$，若 $F(x)$ 的导数 $F'(x)$ 存在，则记为 $f(x)$，并称为是 X 的**密度函数**.

依微积分基本定理有：

$$\int_{-\infty}^{x}f(x)dx=F(x),$$

对相应的离散情形 $p_k = P(X(\omega) = x_k)$ $k = 1, 2, \cdots$ 为 $X(\omega)$ 的离散分布，则分布函数满足：

$$F(x) = \sum_{\{k:x_k \leqslant x\}} p_k.$$

直接讨论分布与密度仍稍嫌繁琐，又引入随机变量的数字特征，几个最常用的量为均值、方差和相关系数等.

定义 随机变量的全部取值依可能性大小作加权平均称为是 $X(\omega)$ 的**数学期望**（或均值），记为 EX. 期望的计算关于离散和连续的分布分别为：

$$EX = \sum_k x_k p_k,$$

$$EX = \int_{-\infty}^{\infty} x f(x) dx.$$

随机变量关于数学期望的平均平方偏差称作**方差**，记为 $\mathrm{Var}X$. $\mathrm{Var}X = E(X - EX)^2$.

例如，对前面碰到的两点分布易证有 $EX = p$，$\mathrm{Var}X = pq$. EX 叫一阶矩，EX^k（$k=1, 2, \cdots$）称为 k 阶矩. 方差是一、二阶矩的函数.

当遇到多个随机变量时则要研究它们之间的相互关系. 最简单的相互关系是互相没有关系即所谓独立性. 与独立相对的概念是相关. 衡量两随机变量相互间线性关系的重要数字特征是**协方差**：$\mathrm{Cov}(X, Y) = EXY - EX \cdot EY$. 其无量纲的标准化即为**相关系数**：

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}X \cdot \mathrm{Var}Y}}.$$

当 X 与 Y 独立时，$\rho(X,Y) = 0$，即 X、Y 不相关. 但反过来一

般不必成立. 与本书讨论主题有关的一个量称为极大相关，这是 Renyi1959年研究的，记为 ρ_M. 其定义为

$$\rho_M(X,Y) = \sup_{\substack{\text{所有Borel} \\ \text{函数} f,g}} \rho(f(X),g(Y)).$$

当 $Y = f(X)$，且 $f(\cdot)$ 为非线性函数，则 $\rho_M = 1$. 但是 $\mathrm{Cov}(X,f(X))$ 却可能是0，这表明不相关所刻画的是没有线性的相关关系，而并不能说明 X 与 Y 之间没有非线性的关联. 常用统计方法多适用于线性的模型，而对后面要讨论的混沌现象中的大量非线性迭代函数常显得贫乏无力.

除研究单个事件的概率外，概率论中还要讨论给定了事件 B 发生的条件下 A 发生的概率称作是**条件概率**，记为 $P(A|B)$. 条件概率满足概率的许多性质，其计算公式为

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

当 $P(A \cap B) = P(A) \cdot P(B)$ 或 $P(A|B) = P(A)$ 时事件 A 和 B 称为是独立的. 并可将此概念推展到多个事件两两独立以及随机变量之间的独立性. 当事件 B 分划成不交事件之并 $\bigcup_{i=1}^{n} B_i$ 时，有所谓全概率公式

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i),$$

它可使计算简化. 有了条件概率便可定义给定一个随机变量 X 时另一随机变量的**条件数学期望**，记作 $E(Y|X)$. 这些内容都已无法一一展开了.

借助均值和方差等数字特征，人们能以一定的概率把握随机变量 X 取值的范围. 一个著名的不等式可以表述为

定理1.1.1 (Chebychev) 记 X 为均值为 μ，方差为 σ^2

的随机变量，则对任给的 $\varepsilon>0$都有

$$P\{|X-\mu|\geqslant\varepsilon\}\leqslant\frac{\sigma^2}{\varepsilon^2}.$$

证明 以离散型随机变量为例，由方差的定义

$$\begin{aligned}\sigma^2&=\sum_k(x_k-\mu)^2p_k\\&=\sum_{k:|x_k-\mu|\geqslant\varepsilon}+\sum_{k:|x_k-\mu|<\varepsilon}\\&\geqslant\sum_{k:|x_k-\mu|\geqslant\varepsilon}(x_k-\mu)^2p_k\\&\geqslant\varepsilon^2P\{|X-\mu|\geqslant\varepsilon\}.\end{aligned}$$

连续情形把和式改为积分即可类似得证. □

Chebychev 不等式虽然估得不细但是很实用.

§1.2 正态分布

在林林总总的概率分布大家庭中最具中心位置的是正态分布. 在离散分布类中在某些方面有类似作用的要数 Poisson 分布. 本节对它们作简略介绍. 正态分布常被称为是 Gauss 分布，其密度呈钟形曲线，如图1.1所示两头小，中间大. 其曲线

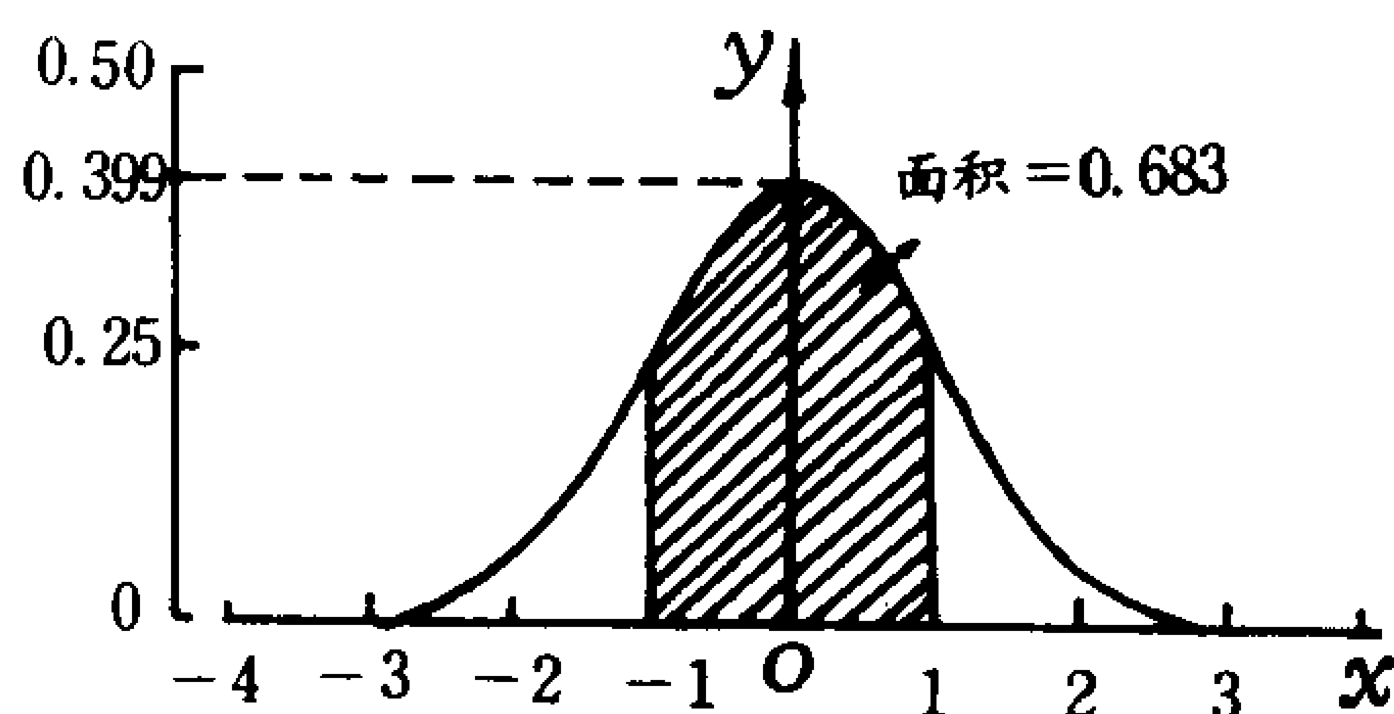


图1.1 标准正态曲线

方程为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

其中 x 在 $(-\infty,\ \infty)$ 上变化，μ 代表均值，曲线关于 μ 对称，σ^2代表方差，σ^2的大小意味着钟的宽狭或曲线平坦的程度．当 x 远离 μ 时 $f(x)$ 很快地趋于0．当 $\mu=0$，$\sigma^2=1$时，以 f 为密度的随机变量 X 称为是服从标准正态分布的，记作 $X\sim N(0,1)$．De Moivre（1667—1754）在计算赌博获胜概率时用正态分布来近似计算二项概率（文章发表的日期是1733年11月12日）．Gauss 则在天文测量误差的研究中发现了它．Galton 称正态分布为“均数偏移律”和“误差频数律”，并惊叹道：“没有什么东西能象宇宙秩序表示为误差频数律这一美妙形式那样令人印象深刻”．如今，多少人对混沌现象和各式各样美妙的分形集发出了同样的惊叹．

在 n 次独立的扔钱币 Bernoulli 试验中，记正面出现的次数为 X，X 服从二项分布．记 $P(X=k)$ 为 $b(k;n,p)$，并记标准正态分布的密度函数为 $\varphi(x)$，分布函数为 $\Phi(x)$，则 de Moivre 得到了

命题1.2.1 记 $x_k=(k-np)\frac{1}{\sqrt{npq}}$，则当 $n\to\infty$，$k\to\infty$ 及$\frac{1}{\sqrt{n}}x_k^3\to0$时，有

$$b(k;n,p)\sim\frac{1}{\sqrt{npq}}\varphi(x_k).$$

证明在普通概率论教本中不难找到，而图1.2则提供了直观的佐证．其中 $P(k)=b\left(k;n,\frac{1}{2}\right)$．

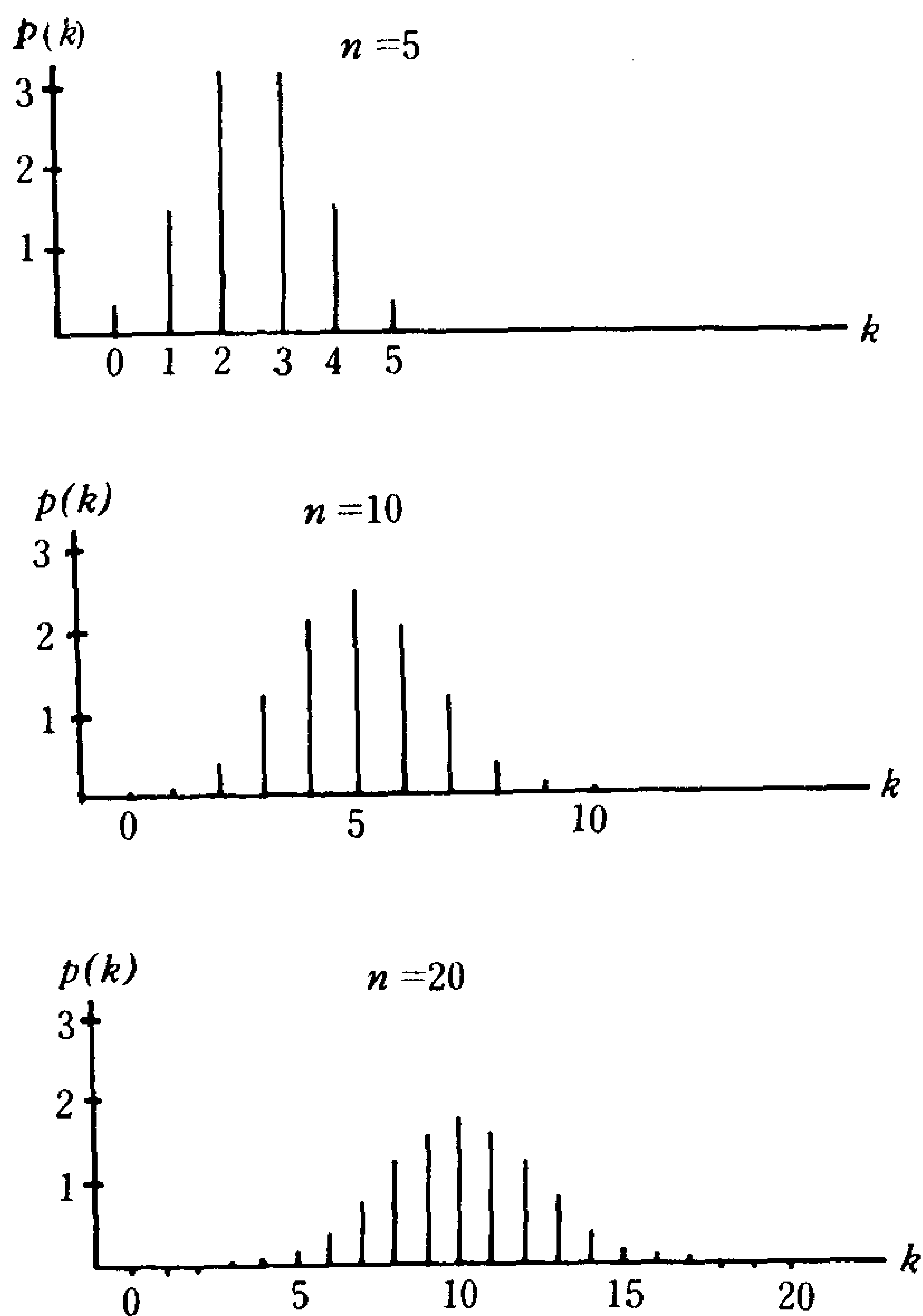


图1.2

正态分布有许多有趣的性质：比如若 x_1，…，x_n 分别服从正态分布则它们的任意线性组合仍然服从正态分布，而且反过来若（X_1，…，X_n）的任何线性组合均服从正态分布，则（X_1，…，X_n）服从多维的正态分布．由此可见 Gauss 分布与线性模型之间深刻的内在联系．在0均值，单位方差的连续型密度中正

态分布是使 Boltzmann 熵 $-\int_{-\infty}^{\infty} f(x)\log f(x)dx$ 达到最大的分布密度，正态分布随处可见与向熵高方向发展的自然规律—热力学第二定律不谋而合.

有些事件，诸如灾害、事故等稀有事件，其发生概率 p 很小但 np 却可能趋于一个定值 λ. 这时二项分布 $b(k,n,p)$ 概率的逼近公式就不是命题1.2.1中的结果而是参数为 λ 的 Poisson 分布，其概率为：

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda},\quad k=0,1,2,\cdots$$

命题1.2.2 当 $n\to\infty$，$np\to\lambda$ 时，

$$b(k;n,p)\sim\frac{\lambda^k}{k!}e^{-\lambda}.$$

证明 注意到

$$b(0;n,p)=\binom{n}{0}p^0q^n\doteq\left(1-\frac{\lambda}{n}\right)^n.$$

当 $n\to\infty$ 由微积分知识知 $b(0;n,p)\sim e^{-\lambda}$. 再利用 $b(k;n,p)/b(k-1;n,p)\sim\lambda/k$ 和归纳法即可得所欲证.

Poisson 分布的均值 $\mu=\lambda$，方差 $\sigma^2=\lambda$ 而且各阶联合半不变量也都等于 λ. 作者在1991年证明了满足这一矩性质的非负整数值随机变量一定是 Poisson 分布. Poisson 分布的最早形式也由 de Moivre 获得，而命题1.2.2则是 Poisson 在1837年证明的. 1898年就有人用 Poisson 分布来拟合骑兵团队中不幸被马踢伤致死的人数的分布，近年更有人把球赛中的射门得分作了 Poisson 拟合，这样使大致预测球赛的结果成为可能. 在有关分形的相关维数的估计中人们也用到了 Poisson 逼近的方法.

§1.3 大数定律与中心极限定理

和扔骰子一样普遍的随机试验是扔钱币. 它甚至成了“随机”的代名词（在§4.1将会看到混沌向扔钱币的挑战）. 正面向上还是反面向上决定了足球队首攻的方向，人们借此询问天意. 这种两点分布的概率模型无处不在，比如下雨这一自然现象是由极多的单个雨滴的落下这种大量的随机事件所构成的. 如果从下第一滴雨开始留心相邻的左右两块面积大小一样的铺路石被打湿的情形，雨滴的行为可以看作是“随机的”. 依次记录下每滴水落地的位置得到序列：

$$LRLLRRRLRLRRLLLR\cdots$$

（L 代表左，R 代表右）. 我们无法预知下一滴将打在左边还是右边，但我们毫不怀疑雨后落在两块石头上的雨点数大体相同. 雨量计的设计就基于这种确信. 局部的细节的不可预知与整体的一定数量规模上的可以预知、可以把握构成了偶然与必然的有机统一. 研究这类规律的理论叫概率论的极限理论，是概率论与数理统计的重要基础. 它的主要成果体现在一系列的大数定律和中心极限定理上. 在大量的随机试验中，经验频率接近于理论频率这一规律称为大数定律. 其最早形式由 Bernoulli 在1713年出版的《推测术》中给出. 依据大数定律，我们可以知道在前10000滴雨水中会有约一半落在左边的石块上. 甚至连误差的大小都可以用概率的语言加以描绘：有95%的把握可以断言左边石块上的雨点数在4900至5100滴之间. 这类大数定律叫关于概率的大数定律，由它可以推出关于均值的大数定律. 记录随机变量 X 的 n 次试验或观察值 $X_1, \cdots, X_n$，则其算术平均

$\overline{X}_n=\frac{1}{n}\sum_i X_i$ 当 n 很大时会接近 X_i 的共同均值 EX. 设 X_i 中取值为 x_k 的有 m_k 个，$k=1$，2，…，于是

$$\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n) = \frac{1}{n}\Big(\sum_k m_k x_k\Big) = \sum\left(\frac{m_k}{n}\right)x_k.$$

由
$$\frac{m_k}{n} \to p_k = P(X = x_k)$$

即可推出

$$\overline{X}_n \to \sum_k x_k p_k = EX.$$

关于均值的大数定律是统计估计的基础. 若引入事件 A 的示性函数$1_A=1$，当 A 发生；$=0$，当 A 不发生. 1_A 也是随机变量，由 $E1_A = P(A)$，关于均值的大数定律蕴含着关于概率的大数定律.

再看一个例子：x 为 [0，1] 区间上的一个实数，其2进位小数展开为$0.a_1a_2a_3\cdots$. 为使展开式唯一，约定 x 的小数表示中无穷多个1相继排列时不进位到有限位. x 既可以看成从 [0，1] 区间上随机地抽取的一个数，也可以看成一连串扔币随机试验的结果，以 a_i 取1代表正面，a_i 取0代表反面. 这时样本空间 Ω 为区间 [0，1]，$\mathscr{B}$ 是 [0，1] 中的 Borel 集全体. 概率 P 为[0,1]中的 Lebesgue 测度. 每个 a_i 还可看成是 x 的函数 $a_i(x)$，并显然有

$$x = \sum_{n=1}^{\infty} \frac{1}{2^n}a_n(x).$$

若 x 的前 n 位二进小数为 x_1，…，x_n，即 $a_i(x) = x_i, i = 1,2,\cdots,n$，则

$$x \in \left[\sum_{i=1}^{n} \frac{x_i}{2^i}, \sum_{i=1}^{n} \frac{x_i}{2^i} + \sum_{i=n+1}^{\infty} \frac{1}{2^i}\right]$$

这类区间叫 n 阶二值区间，由全体前 n 位二进小数为 $x_1\cdots x_n$ 的

x 组成的集合其测度为 n 阶二值区间的长度 $\frac{1}{2^n}=\sum_{i=n+1}^{\infty}\frac{1}{2^i}$，由 $a_i(x)$ 服从两点分布知 $\sum_{i=1}^{n}a_i(x)$ 服从二项分布 $B(n,p)$，其中 $p=\frac{1}{2}$. 有了这些准备，我们即可讨论大数定律的两个特殊形式. 所陈述的内容都是在某种意义下前 n 位二进小数中约有一半是 0，一半是1.

定理1.3.1 （弱大数定律）对任给的 $\varepsilon>0$，

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}a_i(x)-\frac{1}{2}\right|\geqslant\varepsilon\right\}=0.$$

证明 上式的直观解释为当 n 很大时前 n 位小数不是0，1各半的概率将会很小很小. 现往证之. 记 $r_i(x)=2a_i(x)-1$，则 $r_i(x)$ 各以概率 $\frac{1}{2}$ 取值 ± 1(这类变换的手法叫中心化). 命题遂等价于对 $S_n(x)=\sum_{i=1}^{n}r_i(x)$. 证明

$$\lim_{n\to\infty}P\left\{\left|\frac{S_n(x)}{n}\right|\geqslant\varepsilon\right\}=0.$$

因为每个 $i-1$阶二值区间都分成两个 i 阶二值区间，且 $r_i(x)$ 在左、右两区间各取值 -1 和 $+1$，所以对所有 $r_i(x)$ 有

$$\int_0^1 r_i(x)dx=0,\qquad \int_0^1 r_i^2(x)dx=1.$$

又容易看出对 $i<j\leqslant n$ 有

$$\int_0^1 r_i(x)r_j(x)dx=0.$$

也即 r_i，r_j 为不相关的随机变量. 于是

$$ES_n(x)=\sum_i Er_i(x)=0,$$

$$\mathrm{Var}S_n(x) = \sum_i \mathrm{Var}(r_i(x)) = n.$$

再由 Chebychev 不等式，当 $n\to\infty$，

$$P\left\{\left|\frac{S_n(x)}{n}\right| \geqslant \varepsilon\right\} = P\{|S_n(x)| \geqslant n\varepsilon\}$$

$$\leqslant \frac{\mathrm{Var}S_n(x)}{n^2\varepsilon^2} = \frac{n}{n^2\varepsilon^2} = \frac{1}{n\varepsilon^2} \to 0. \quad \square$$

一个更强的结果是

定理1.3.2 （强大数定律）记 N 为集合 $\left\{x \,\middle|\, \lim\limits_{n\to\infty} \frac{S_n(x)}{n} = 0\right\}$，则 $P(N) = 1$.

证明 由 Chebychev 不等式的推广形式有

$$P\{|S_n(x)| \geqslant n\varepsilon\} \leqslant \frac{1}{n^4\varepsilon^4}\int_0^1 S_n^4(x)dx.$$

再注意到

$$r_i^4(x) = 1,$$

$$r_i^2(x)r_j^2(x) = 1,$$

$$r_i^2(x)r_j(x)r_k(x) = r_j(x)r_k(x),$$

$$r_i^3(x)r_j(x) = r_i(x)r_j(x),$$

及对不同的 i，j，k，l 有

$$\int_0^1 r_i(x)r_j(x)r_k(x)r_l(x)dx = 0,$$

则容易估计出

$$\int_0^1 S_n^4(x)dx = n + 3n(n-1) \leqslant 3n^2.$$

所以

$$P\left\{\left|\frac{S_n(x)}{n}\right| \geqslant \varepsilon\right\} \leqslant \frac{3n^2}{n^4\varepsilon^4} = \frac{3}{n^2\varepsilon^4}.$$

取 $\varepsilon_n = \frac{1}{\sqrt[8]{n}}$. 并记 $A_n = \left\{x \middle| \left|\frac{S_n(x)}{n}\right| \geqslant \frac{1}{\sqrt[8]{n}}\right\}$,则

$$P\{A_n\} \leqslant \frac{3}{n^2\sqrt{n}}.$$

于是 $\sum_n P(A_n) < \infty$. 作为收敛级数的尾部,对任给的 $\varepsilon > 0$,又有 $\sum_{n\geqslant m} P(A_n) \leqslant \varepsilon$. 记 N 的补集为 N^C. N^C 含在集 $\bigcup_{n\geqslant M}^{\infty} A_n$ 之中,故有

$$P\{N^C\} \leqslant P\{\bigcup_{n\geqslant m} A_n\} \leqslant \sum_{n\geqslant m} P\{A_n\} \leqslant \varepsilon.$$

由 ε 的任意性知

$$P\{N^C\} = 0,$$

$$P\{N\} = 1 - P\{N^C\} = 1. \quad \square$$

如果 N^C 为可数集则其测度为 0，但这里的 N^C 却并非可数集. 有兴趣的读者可以想一想为什么？强大数定律告诉我们当 n 很大之后对 [0，1] 中几乎所有的 x 其前 n 位二进小数中都有一半是 0，一半是 1. 这也是著名的 Borel 正则数定理. N 中的 x 就称为正则数，[0，1] 中几乎所有的数都是正则数. 这一结果是 Borel 在 1909 年发表的. 这两个大数定律的证明尽显古典大数定律的精华. 条件不多但结论深刻. 推而广之，Borel 正则数定理还告诉人们在 [0,1] 中的几乎所有的十进制小数 $0.x_1x_2\cdots x_n\cdots$ 当 n 很大时前 n 位小数中数码 0,1,…,9 将大约各占 $\frac{1}{10}$. 此时读者该不会对这一结论感到惊奇了.

强大数定律和弱大数定律是因其不同的收敛模式而得名的. 以 X_n 趋于 0 为例，强收敛即以概率为 1 收敛要求 $P\{\lim X_n = 0\} = 1$;弱收敛即依概率收敛要求 $P\{|X_n| \geqslant \varepsilon\} \to 0$. 前者显然强于后者. 一般形式的大数定律可表述为

定理 1.3.3 (Chebychev) 若 $\{X_n\}$ 为一串独立随机变量，它们共同的均值 μ 和方差 σ^2 有穷，则 $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ 依概率收敛到 μ.

定理 1.3.4 (Kolmogorov 强大数定律) 若 $\{X_n\}$ 为一串独立同分布（以后记为 i, i, d）的随机变量序列，则 $\overline{X}_n$ 以概率为 1 收敛于一个有限常数 C 的充分必要条件是 $E|X_1|<\infty$ 和 $EX_1=C$.

定理 1.3.4 的标准证法要利用随机变量截尾的技巧和 Kolmogrov 准则：若 $\{X_i\}$ 独立，零均值，方差有穷，又对某一串趋于无穷的正常数 $\{b_n\}$ 有

$$\sum_{n=1}^{\infty} EX_n^2/b_n^2 < \infty,$$

则有

$$\sum_{i=1}^{n} X_i/b_n \to 0.$$

以概率为 1 成立. 对方差不必有穷的一般随机变量常将 X_i 截尾为

$$X_i^c = \begin{cases} X_i, & \text{当 } |X_i| \leqslant b_i; \\ 0, & |X_i| > b_i. \end{cases}$$

记 $\tilde{X}_i = X_i^c - EX_i^c$ 为 X_i^c 的中心化. 则当

$$\sum_{n=1}^{\infty} P\{|X_n| > b_n\} < \infty$$

和

$$\sum_{n=1}^{\infty} \frac{1}{b_n^2} E\tilde{X}_n^2 < \infty$$

时有

$$\lim_{n\to\infty} \frac{1}{b_n} \sum_{i=1}^{n} X_i = 0.$$

数学家们还对定理的条件作了各种可能的推广. 放宽 $\{X_n\}$ 独立的条件为一般的平稳序列和鞅序列. 在各种相依条件下也有相应的结果. 每年与极限定理有关的概率统计学术论文不下数百篇. 强大数定律还是 § 2.4 中一般遍历性定理的一个最特殊的情形. 在关于混沌的数学定义中混合性的概念与遍历是相通的，在这个意义上随机性与混沌现象也是相通的. 对二者进行区分反而成了统计学的重要课题.

大数定律在统计估计的理论中举足轻重. 人们知道 n 次 Bernoulli 试验的平均成功次数是成功概率的一个很不错的估计. 而衡量估计量好坏的重要标准是相合性：估计量以某种方式趋近于被估计量. 而大数定律则是相合概念的基础. 在统计学中，观察数据的母体的分布 $F(x)$ 是未知的，大量的分析建立在经验分布之上. 基于随机的观察样本 $X_1, \cdots, X_n$ 经验分布的定义为：

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leqslant x\}}.$$

也即样本中 $\leqslant x$ 的样本个数与总样本数的比值. 而 $1_{\{X_i \leqslant x\}}$ 又是两点分布，成功概率为 $p = F(x)$. 于是大数定律保证了 $F_n(x)$ 将以某种方式趋于 $F(x)$. 其重要性不言而喻.

极限定理的另一主题是中心极限定理. 其要旨是讲一些实际自然现象若受到许多相互独立的随机因素的影响，且每个因素的影响都很小时，总的影响可以看作是服从正态分布的. 中心极限定理的名称是著名数学家 Polya 在 1920 年起的，以强调这类定理在概率论和数理统计中的“中心”作用. 1935 年左右经过 de Moivre, Laplace, Levy, Feller, Lyapunov, Lindeberg 等一批数学家的努力，终于有了完满的结果.

考查一串独立的随机变量序列 $\{X_n\}$.

设 $EX_k=\mu_k$, $\mathrm{Var}X_k=\sigma_k^2$, $S_n=\sum_{k=1}^{n}X_k$ 称为 $X_1,\cdots,X_n$ 的部分和，$s_n=\sum_{k=1}^{n}\sigma_k^2$ 是 $X_1,\cdots,X_n$ 方差的和，把 S_n 标准化(使其均值为 0，方差为 1) 为

$$T_n=\frac{S_n-\sum_k \mu_k}{s_n}$$

则有

定理 1.3.5 在上述记号下 $\lim_{n\to\infty}P\{T_n\leqslant x\}=\Phi(x)$，且 $\lim_{n\to\infty}\max_{1\leqslant k\leqslant n}\sigma_k^2/s_n^2=0$ 的充分必要条件为 Lindeberg 条件

$$\lim_{n\to\infty}\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x-\mu_k|\geqslant \tau s_n}(x-\mu_k)^2dF_k(x)=0.$$

其中 $F_k(x)$ 为 X_k 的分布函数，τ 为任给的正实数.

中心极限定理的证明方法主要有特征函数的方法和矩的方法. 前者依赖随机变量的 Fourier 变换；后者利用随机变量相应矩的收敛而导出分布函数的收敛. 中心极限定理就所阐明的内容而言大体有三类：上述的定理 1.3.5 主要处理的是单个序列的部分和在标准化后的极限分布；第二类是考虑更一般的三角阵列 $\{X_{n,j}: j=1, 2, \cdots, k_n, n=1, 2, \cdots\}$ 所考虑的是和 $S_n=\sum_j X_{nj}$ 的极限分布；第三类应属于逼近定理. 它研究和估计部分和分布与其逼近分布之间的距离，最著名的是 Berry-Essen 定理. 就本书的主题而言，我们将更多地讨论遍历性，仅在布朗运动的讨论中涉及中心极限定理，本章的命题 1.2.1 讨论的是局部极限定理，是讲逼近正态分布的密度函数的. 极限理论

内容浩繁，结论深刻，是概率统计的重要研究领域.

§1.4 随机游动与反正弦律

大数定律建立起了偶然与必然之间的桥梁. 人们满有把握地说当扔钱币的次数足够多时，将有近一半是正面（如果钱币是均匀的）. 于是扔钱币也成为公平的象征；甚至代表上帝的旨意. 这么明显的凭直观即普遍乐意接受的结论似乎并不需要化很大的力气来论证一番. 但如若另看一个例子，直观有时也会引人误入歧途：甲乙二人每秒钟扔一次钱币进行赌博，胜者赢一块钱. 某甲连续玩了数百万次，更精确地说是 362 天 15 小时而一直保持着领先权. 所谓领先是指累计赢利为正. 这件事的概率有多大呢？微乎其微吧！不然，此事的概率为 0.05，每 20 人中就有一个象某甲这样的幸运儿. 比商场购物的幸运大奖的中奖率要高出百倍. 而这也恰恰是一个正常钱币所表现出来的必然的规律性. 请看论证.

记每次掷币的结果为随机变量 X_n，X_n 取值 ± 1，概率各为 $\frac{1}{2}$. $S_n=\sum_{i=1}^{n} X_i$ 则是 n 次赌局之后的累积赢利. S_n 也产生于质点从原点出发，各以 $\frac{1}{2}$ 概率向左或向右移动一个单位距离的模型. S_n 代表时刻 n 质点所处的位置，$\{S_n, n\geqslant 0\}$ 为一对称的随机游动. 记在时刻 $2n$ 甲乙玩成平局的概率为 u_n. 则容易算出

$$u_n = P\{S_{2n}=0\} = P\{2n \text{ 次独立试验中成功 } n \text{ 次}\}$$
$$= \binom{2n}{n}\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^n = \binom{2n}{n}\frac{1}{2^{2n}}.$$

引理 沿用上述记号，则

$$P\{S_i \neq 0, i = 1, \cdots, 2n\} = u_n.$$

证明 由 $P\{S_1 \neq 0, S_2 = 0\} + P\{S_1 \neq 0, S_2 \neq 0\} = 1$ 知 $P\{S_1 \neq 0, S_2 \neq 0\} = 1 - u_1$，而在 $2n$ 时首次出现平局的概率 $P\{S_1 \neq 0, S_2 \neq 0, \cdots, S_{2n-1} \neq 0, S_{2n} = 0\}$ 依古典概率计算选票问题的结果知为$\frac{u_n}{2n-1}$. 于是由

$$\begin{aligned} P\{\Omega\} &= P\{S_1 \neq 0, S_2 = 0\} + P\{S_1 \neq 0, S_2 \neq 0\} \\ &= P\{S_1 \neq 0, S_2 = 0\} \\ &\quad + P\{S_1 \neq 0, S_2 \neq 0, S_3 \neq 0, S_4 \neq 0\} \\ &\quad + P\{S_1 \neq 0, S_2 \neq 0, S_3 \neq 0, S_4 = 0\} = \cdots \end{aligned}$$

知 $$P\{S_i \neq 0, i = 1, \cdots, 2n\} = 1 - \sum_{k=1}^{n} \frac{u_k}{2k-1}.$$

而由归纳法不难证明 $u_n = 1 - \sum_{k=1}^{n} \frac{u_k}{2k-1}$.

(注意 $u_1 = \frac{1}{2}$ 及 u_n 的递推关系 $2nu_n = (2n-1)u_{n-1}$).

记对称随机游动在 $2n$ 时曾有 $2k$ 个时间单位为正，$2n-2k$ 个时间单位为负这一事件为 $E_{n,k}$. 并记 $P(E_{n,k}) = P_{n,k}$，则有

定理 1.4.1 $P_{n,k} = u_k u_{n-k}$.

证明 由 $u_0 = 1$，$u_1 = \frac{1}{2}$ 而 $P_{1,0} = P_{1,1} = \frac{1}{2}$，故当 $n=1$ 时命题成立.

记从 0 出发首次回到 0 的时刻为 T，T 也是随机变量. 当 $X_1 + X_2 = 0$ 时，T 取值 2. 由全概率公式

$$\begin{aligned} P_{n,n} = &\sum_{i=1}^{n} P\{E_{n,n} | T = 2i\} P\{T = 2i\} \\ &+ P\{E_{n,n} | T > 2n\} P\{T > 2n\}. \end{aligned}$$

由于从正、负两侧回到 0 的方式有对称性，

$$P\{E_{n,n}|T=2i\}=\frac{1}{2}P_{n-i,n-i}$$

$$P\{E_{n,n}|T>2n\}=\frac{1}{2}.$$

由归纳法可假定对所有 $m<n$，有 $P_{m,k}=u_k u_{m-k}$，所以

$$\begin{aligned}
P_{n,n} &= \frac{1}{2}\sum_{i=1}^{n}P_{n-i,n-i}P\{T=2i\}+\frac{1}{2}P\{T>2n\} \\
&= \frac{1}{2}\sum_{i=1}^{n}u_{n-i}u_0P\{T=2i\}+\frac{1}{2}P\{T>2n\} \\
&= \frac{1}{2}\sum_{i=1}^{n}P\{S_{2n-2i}=0\}P\{T=2i\}+\frac{1}{2}P\{T>2n\} \\
&= \frac{1}{2}\sum_{i=1}^{n}P\{S_{2n}=0|T=2i\}P\{T=2i\} \\
&\quad +\frac{1}{2}P\{S_1\neq 0,\cdots,S_{2n}\neq 0\} \\
&= \frac{1}{2}P\{S_{2n}=0\}+\frac{1}{2}u_n=u_n.
\end{aligned}$$

再由正与负的对称性知 $P_{n,0}=P_{n,n}=u_n$. 对 $0<k<n$ 可用归纳法及类似的条件化手法证之.

一个随机变量取值 $2k$ 的概率为 $u_k u_{n-k}$则称作是离散的反正弦分布. 定理 1.4.1 表明在 $2n$ 次独立扔币中正面出现 $2k$ 次的概率服从反正弦律. 取 $n=10$，算出

$$P_{20,0}=0.1762>P_{10,10}=0.0606.$$

可见一边倒的概率为势均力敌概率的3倍. 由此理解中日围棋擂台赛中某方五连胜就不足为奇了. 此概率分布的概率直方图形如"∪"字. 当n很大时由著名的 Stirling 公式 $n!\sim\sqrt{2\pi}n^{\frac{n+1}{2}}e^{-n}$知

$$P_{2n,2k} \sim \frac{1}{\pi \sqrt{k(n-k)}}.$$

借助数学分析中黎曼和向定积分的收敛有

定理 1.4.2 对固定的 $0<\alpha<1$，S_n 处于正的时间与总时间长度之比 $\frac{k}{n}<\alpha$ 的概率当 $n\to\infty$ 时趋于

$$\frac{1}{\pi}\int_0^\alpha \frac{dx}{\sqrt{x(1-x)}} = \frac{2}{\pi}\arcsin\sqrt{\alpha}.$$

反正弦律因此得名. 在上式中取 $\alpha=\frac{2.24\text{天}}{365\text{天}}\doteq 0.006137$，则 $\frac{2}{\pi}\arcsin\sqrt{\alpha}\doteq 0.05$，这就解释了“好运气”这一必然规律. 定义 $S_0=0$. 在随机游动 $\{S_n, n\geqslant 0\}$ 中如果有 k 使 $S_k>0$ 及当 $i<k$ 时 $S_k>S_i$ 和 $j>k$ 时 $S_k>S_j$，则称 k 为最大值 S_k 所处的位置. 当对任何 $n\geqslant 1$ 时有 $S_n\leqslant 0$，称 0 为最大值的位置. 于是又有

定理 1.4.3 在 $2n$ 次随机游动中，最大值的位置是在 M 的概率记为 P_M，则当 $M\neq 0$，$P_M=\frac{1}{2}u_{2n-2M}u_{2M}$；当 $M=0$，$P_0=u_{2n}$.

这称为第二反正弦律. 它表明最大值位置在两端的概率要大于中间，要么一路顺风，要么倒霉透顶. 本世纪七、八十年代在混沌现象的研究中，人们对简单 logistic 迭代函数 $f(x)=4x(1-x)$ 进行了透彻的研究. 发现此映射所对应的不变概率分布正是反正弦律. 简而言之，若 X 是 $[0, 1]$ 区间上的随机变量，它服从反正弦律则 $f(X)=4X(1-X)$ 也服从反正弦律. 读者不难自行证明，也可在 § 3.1 中找到原由.

第二章　把握随机性

随机性无处不在！

——Howell Tong

Randomness is ubiguitous!

随机性随处可见．概率论也并没有停留在扔钱币和组合计数的古典阶段．到本世纪 30 年代以 Markov 研究马氏过程和 Wiener 研究布朗运动为起点有关随机过程的研究蓬勃发展．随机过程是一切动态的随时间推进（也可以是空间的扩展）而变化着的随机现象的数学抽象．比如，每天的气温、分子受碰撞后运动、电话总机中通话次数、生物群体的盛衰规律、证券市场上的价格指数等等都是随机过程的实现．本章仅简略介绍这一丰富理论中与混沌研究直接有关的遍历性，Markov 链、布朗运动以及时间序列等概念．随机过程及其统计推断这一充满活力、应用广泛的数学分支正是人们深刻地把握随机性的重要工具．

§ 2.1 随机过程和时间序列

随机过程英文叫 Stochastic Process 或 random process. stochastic 一词源于希腊文 στοχαστυκοξ，本意是“猜”，但可绝不是随便乱猜. 各类随机过程的研究都从不同侧面把握着随机现象的规律. 先给出

定义 一族随机变量 $\{X_t,\ t\in T\}$ 就是随机过程，其中 T 是指标集，t 为参数.

t 一般代表时间，当 T 为整数集时 X_n 也称作是随机序列. $T=R$（实数集）时就是连续时间的过程. X_t 是定义在某个概率空间 Ω 之上的故可记为 $X(t,\omega)$，取定随机试验的结果 ω_0，$X(t,\omega_0)$ 就是一条样本路径也是我们所观察的现实轨迹. 若固定时刻 $t=t_0$ 则 $X(t_0,\omega)$ 为一通常的随机变量. 要了解随机过程的概率性质就要了解它们的有限维分布族（这是分布函数概念的自然推广). 对 T 中的任意 n 个时刻 $t_1,\cdots,t_n$ 及实数 $x_1,\cdots,x_n$ 概率

$$P\{X(t_i)\leqslant x_i, i=1,\cdots,n\}=F_{t_1,\cdots,t_n}(x_1,\cdots,x_n)$$

便是有限维分布族. 相应的数字特征为均值函数 $m(t)=E[X(t)]$ 及衡量过程前后不同时刻间相互依赖程度的协方差函数 $r(s,t)=\mathrm{Cov}(X(s),X(t))$.

例 独立同分布的随机序列为最简单的随机过程.

例 （Poisson 过程）一个整数值随机过程 $\{N(t),t\geqslant 0\}$ 满

足下述三个条件就称作强度为 λ 的 Poisson 过程. (i) $N(0)=0$; (ii) $N(t)$ 有独立增量; (iii) 对任何 $t>0$, $s\geqslant 0$. 增量 $N(S+t)-N(s)$ 服从参数为 λt 的 Poisson 分布.

例 (Brown 运动) 见 § 2.2.

随机过程可按其性质进行各种分类. 其中重要的有平稳过程,独立增量过程和 Markov 过程. 先谈过程的平稳性,如果两个随机变量 X、Y 的分布相同则记为 $X\overset{d}{=}Y$. 类似可定义随机向量的等分布.

定义 若对任意的 $t_1,\cdots,t_n\in T$ 和任意的时间间隔 h 都有

$$(X(t_1+h),\cdots,X(t_n+h))\overset{d}{=}(X(t_1),\cdots,X(t_n))$$

则称 $X(t)$ 是严格平稳的.

这时会有 $m(t)=m$ 和 $r(t,s)$ 只与间隔 $t-s$ 有关可记为 $r(\tau)$. $r(0)$ 为方差, $\rho(\tau)=r(\tau)/r(0)$ 称为自相关函数. 仅满足上述两点的过程又称为宽平稳过程. 在通信信号过程中大量地遇到这类过程. 若在互不相交区间上的增量是相互独立的,则过程又称作是独立增量过程. Poisson 过程和 Brown 运动都是这类过程. 它表示在不同区间上过程的统计性质是互不相关的. 随机过程还可依 $X(t)$ 的分布形式而分类定名,比如 Poisson 过程,Gauss 过程等. 由独立序列向相依情形迈一小步即可引入 Markov 链的概念. 最简单的离散时间 Markov 链的取值的状态空间为 $\{0,1,2,\cdots\}$,于是有

定义 如果对任一列状态 $i_0,\cdots,i_{n-1},i,j$ 及对任何 n,随

机过程 $\{X_n, n \geqslant 0\}$ 满足 Markov 性质

$$P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0\}$$
$$= P\{X_{n+1} = j | X_n = i\},$$

则称 X_n 为一离散时间 Markov 链.

Markov 性质的概率直观是给定了当前 (X_n)，未来与过去是统计上独立的. 当上式与时刻 n 无关时又称 X_n 为带平稳转移概率的 Markov 链. 其一步转移概率记为 P_{ij}. 以 P_{ij}为元素的无穷维随机矩阵 $[P_{ij}]$ 记作 $\boldsymbol{P}$，初始分布 $P_i = P(X_0 = i)$ 及 $\boldsymbol{P}$ 完全刻画了一个 Markov 链的演化过程.

例 直线上的非对称随机游动就是以 $\boldsymbol{P}$ 为转移概率矩阵的 Markov 链，其中

$$\boldsymbol{P} = \begin{Vmatrix} 0 & p & 0 & 0 & \cdots \\ q & 0 & p & 0 & \cdots \\ 0 & q & 0 & p & \cdots \\ & & \vdots & & \end{Vmatrix}$$

为了研究 Markov 链，人们引进了一系列的概念：如从状态 i 出发能够在某 n 步到达 j 则称 i 可达 j；互相可达称为互达；若一个 Markov 链的各个状态都互达就称这个链是不可约的. 又使 n 步转移概率 $P_{ii}^{(n)}$（为矩阵 $\boldsymbol{P}$ n 次幂的第 i 个对角元素）大于 0 的所有 n 的最大公约数为 1，则状态 i 称为是非周期的；如果从状态 i 出发最终以概率 1 能返回状态 i，则称 i 是常返的；且当从状态 i 出发平均返回状态 i 的时间为有限时，状态 i 又称为是正常返的. 一个正常返，非周期的状态称为是遍历的 (ergodic). 一个与本书议题有关的重要命题是

定理 2.1.1 对 Markov 链的遍历状态有

(i) $\lim\limits_{n\to\infty} P_{ii}^{(n)} = \dfrac{1}{\mu_i}$，其中 μ_i 是平均常返时间；

(ii) 若 j 可达 i 亦有$\lim\limits_{n\to\infty} P_{ji}^{(n)} = \dfrac{1}{\mu_i}$；

(iii) 若一个不可约 Markov 链中所有状态都是遍历的，则 $\dfrac{1}{\mu_i} = \pi_i$为一平稳分布，它满足 $\sum\limits_i \pi_i = 1, \pi_j = \sum\limits_i \pi_i P_{ij}$；

(iv) 若不可约 Markov 链存在一个平稳分布则这个 Markov 链的所有状态都是遍历且 n 步转移概率的极限分布就是此平稳分布．也即

$$\lim_{n\to\infty} P_{ij}^{(n)} = \pi_j.$$

定理的证明在一般随机过程的教本中不难找到．其要点是基于著名的 Chapman-Kolmogrov 方程 $P_{ij}^{(n)} = \sum\limits_{k=0}^{\infty} P_{ik}^{(n-1)} P_{kj}$ 做无穷小分析．借助定理可以解方程 $\pi_j = \sum\limits_i \pi_i P_{ij}$ 来求出极限分布 $\{\pi_i\}$．有趣的是若 Markov 链以这个平稳分布作为初始分布（即 $P\{X_0=i\} = \pi_i$），则 Markov 链在各时刻的分布都是不变的！而且都等于 $\{\pi_i\}$．将定理中的结论稍加分析，还可以看到过程经过一段时间的演化后，将处于状态 i 的概率与过程在时间长河中，访问状态 i 的次数与总转移步数之比是一样的．也就是说关于时间的平均与关于状态相位的统计平均是一致的．这便是遍历性理论的重要内涵．记 $I_n = 1_{\{X_n=1\}}$，则$\dfrac{1}{m}\sum\limits_{n=0}^{m-1} I_n$ 代表在 m 步过程的转移中访问 i 的平均次数．设初始状态为 $X_0=j$，则这一平均数的条件期望为

$$E\left\{\frac{1}{m}\sum_{n=0}^{m-1}I_n|X_0=j\right\}=\frac{1}{m}\sum_{n=0}^{m-1}P\{X_n=i|X_0=j\}$$

$$=\frac{1}{m}\sum_{n=0}^{m-1}P_{ji}^{(n)}.$$

当 $m\to\infty$，由算术平均与非负序列本身有相同的极限知 $\lim\limits_{m\to+\infty}\frac{1}{m}\sum\limits_{n=0}^{m-1}P_{ji}^{(n)}=\lim\limits_{n\to\infty}P_{ji}^{(n)}=\pi_i$. 定理也因其重要性而称为 Markov 链的基本极限定理.

下面再谈谈时间序列.

随机过程 $X(t)$ 中的参数 t 一般看作时间，人们对在一定时间间隔所得的观察进行研究. 不失一般性又可取时间为单位区间则 t 属整数集 $\mathbf{Z}$.

定义 一族实值随机变量 $\{X_t,t\in Z\}$ 称为是一个时间序列.

若 X_t 的统计性质有时间齐性，则如定义过程平稳性一样可定义平稳时间序列. 这时 $\rho(k)$ 称为 X_t 的自相关函数，记作 ACF. 基于 ACF 对时间序列进行统计分析称为时间域的分析. ACF 的形状和截断点可提示建模的类型及定阶. 另一领域为频率域，它通过对 $\rho(\tau)$ 的 Fourier 变换获得谱密度 $f(\omega)$. ω 为序列变化的频率，在不同频率上对时间序列的变差有不同的贡献. 这种贡献可由谱密度刻画. 时间和频率两域都含有序列的大量信息. 分析时间序列一般要经过数据采集、初步处理、建立模型、模型辨识、模型参数估计、模型诊断和预测这样几个步骤. 最粗略的模型是将序列分解为四部分：T 代表趋势，S 代表季节效应，R 代表误差，C 为未知的隐含周期. 于是当在各部分可加时有

$$X = T + C + S + R.$$

目前已有各种技术来分解出 T 与 S，也可用周期图和谱分析方法来发现 C. 简而言之，时间序列分析是把观察数据中的确定性部分与随机误差部分分离，并对满足宽平稳条件的序列建立起线性的(或者可以化为线性的)模型. 如图 2.1 所示：a)为原始数据，b)为剩余部分.

例 (白噪声)一串独立同分布的随机变量 ε_t，$E\varepsilon_t=0$，$\mathrm{Var}\varepsilon_t=\sigma^2$，称为白噪声序列. 特别当 $\varepsilon_t \sim N(0, \sigma^2)$ 称为是 Gauss 白噪声. 于是 $\rho(0)=1$，$\rho(k)=0$，$k=1, 2, \cdots$，又由公式 $f(\omega)=\frac{1}{\pi}\left[1+2\sum_{k>0}\rho(k)\cos\omega k\right]$ 算出 $f(\omega)=\frac{1}{\pi}$. 当 $\omega \in [0,\pi]$，为

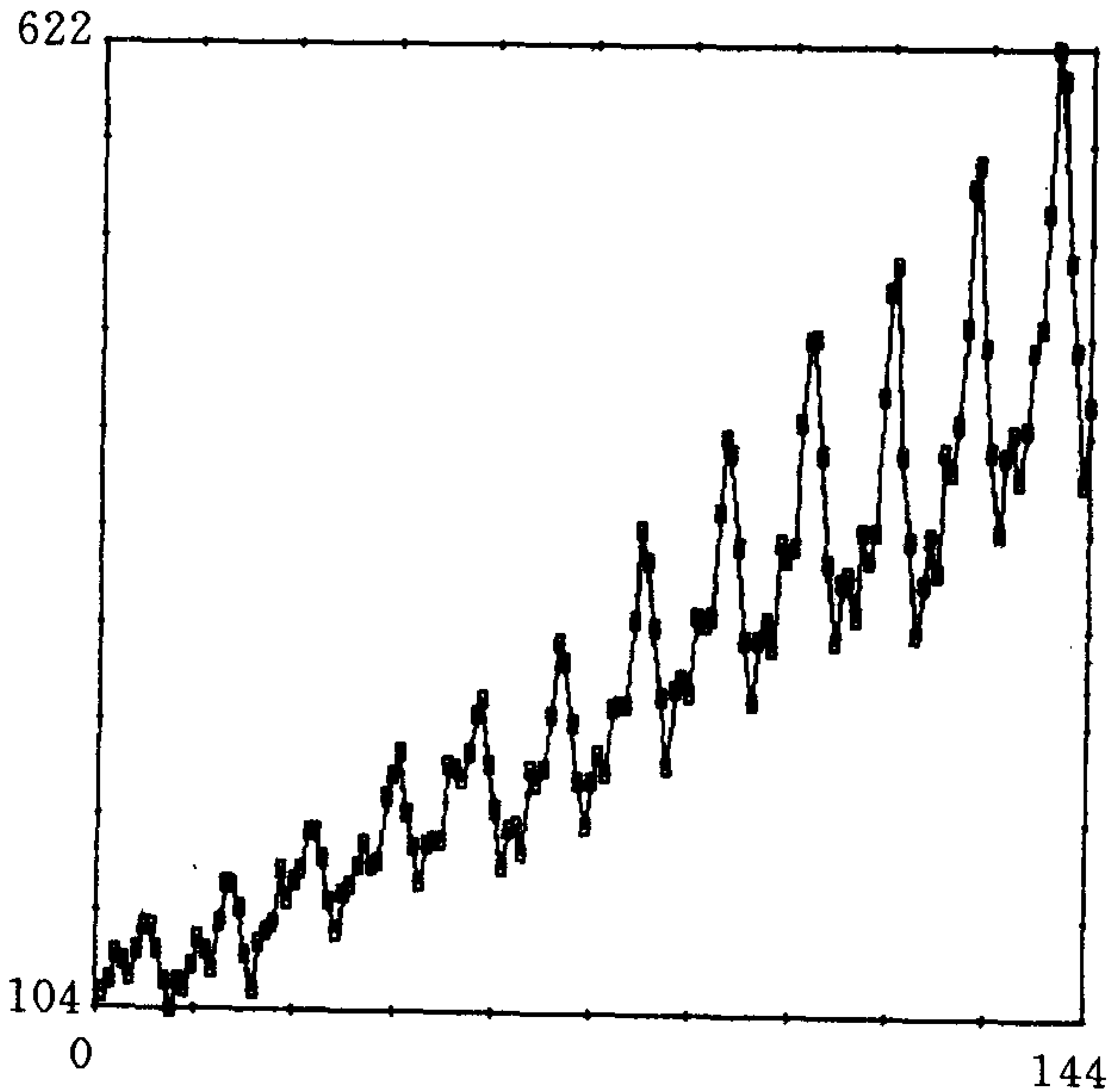


图 2.1 (a)

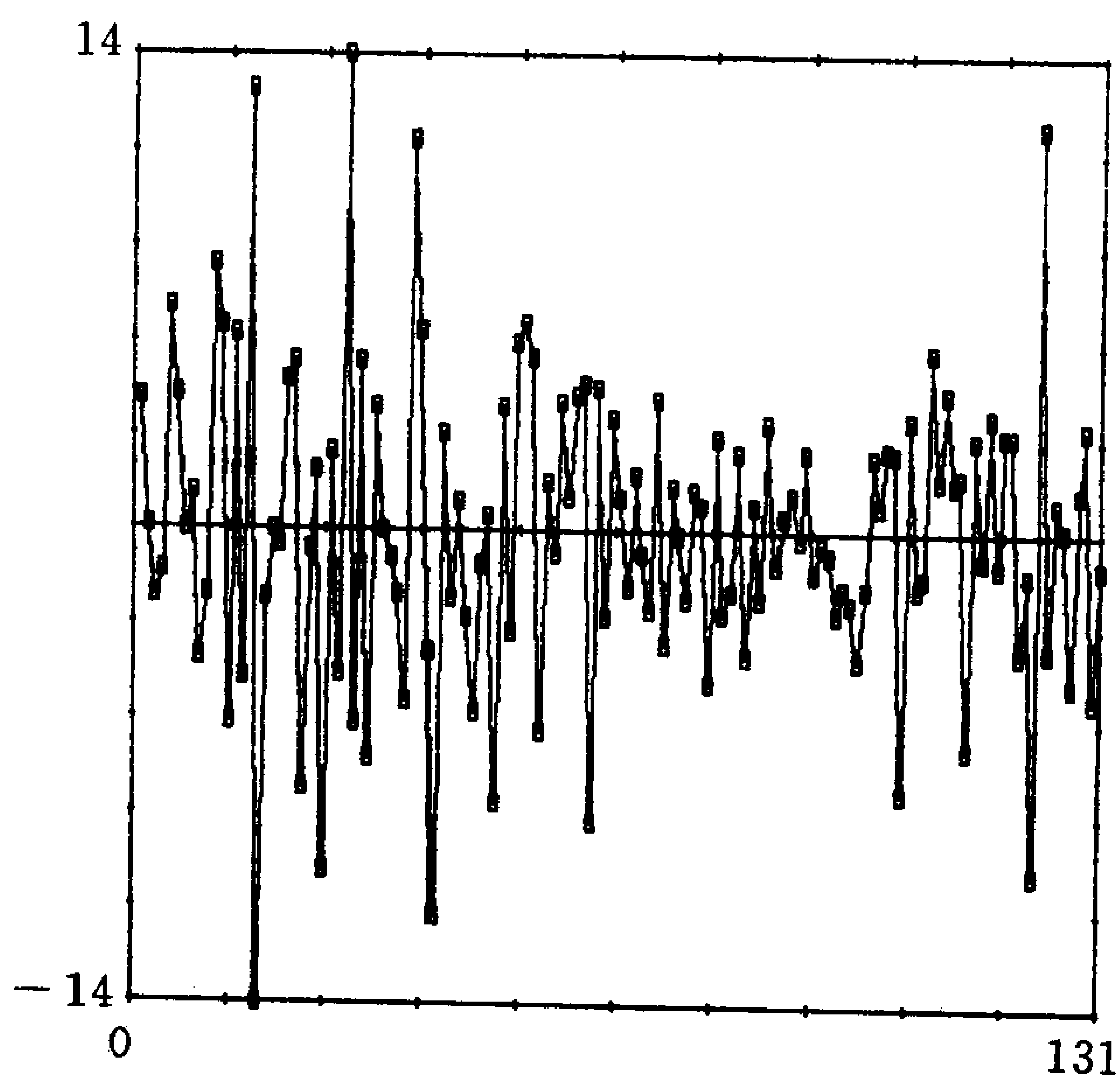


图 2.1 (b)

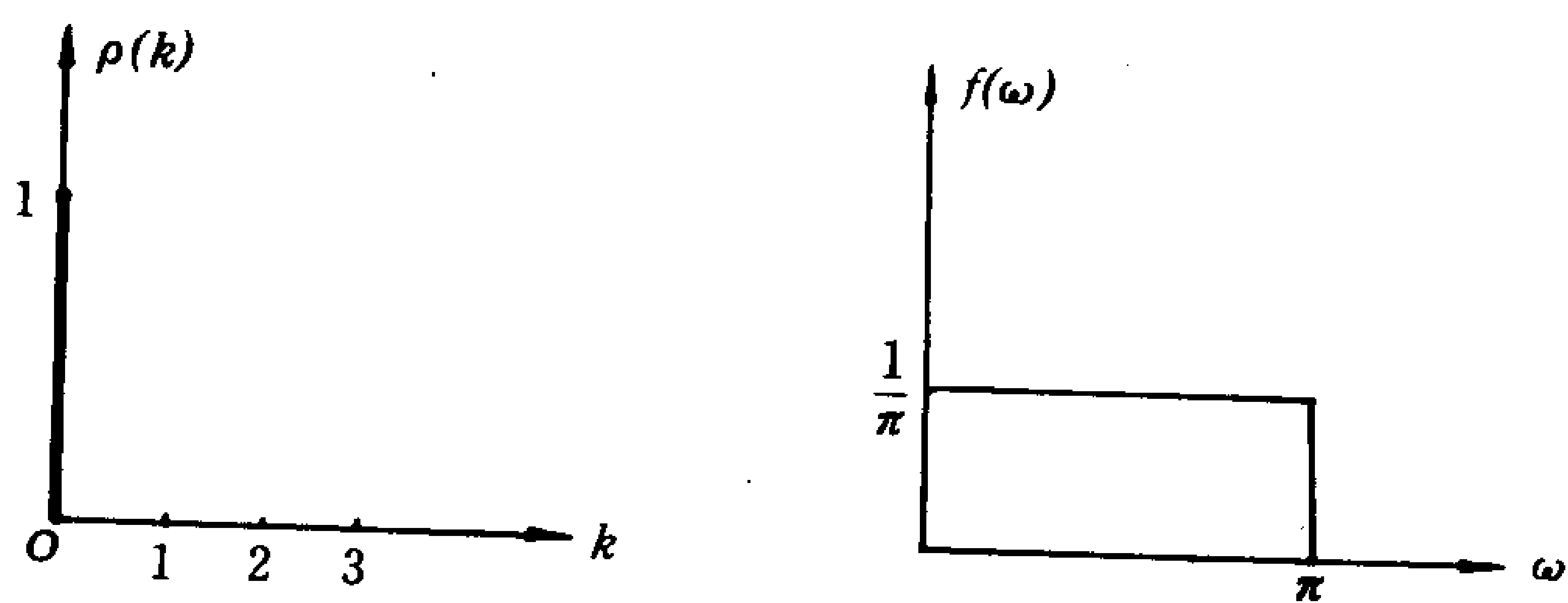


图 2.2

$[0,\pi]$上的均匀分布如图 2.2.

在第三章中，由简单非线函数迭代所产生的混沌现象与随机的白噪声序列，在表面很难用现有的线性统计模型的分析方法加以区别. 区分二者是对当代统计方法的挑战.

Yule 在 1927 年研究太阳黑子数据时已用到了自回归模型. 到本世纪 30 年代 Birkhoff 的遍历性定理及 40 年代 Wold 关于平稳时间序列的非确定性部分，都可以用滑动平均模型来表示的结果铺平了通向近代时间序列分析的道路. 之后，基于白噪声的滑动平均和借用统计中的回归的思想，Box 和 Jenkins 发展了自回归滑动平均混合模型的时间域研究. 为了使数据平滑可先进行差分 $\nabla X_t = X_t - X_{t-1}$，再对 ∇X_t 序列建模.

例 （ARIMA） 一个时间序列 X_t 称为是服从 ARIMA（p，d，q）模型的. 若 $\nabla^d X_t$ 满足

$$\phi(B)\nabla^d X_t = \theta(B)\varepsilon_t,$$

其中 B 为后移算子，$BY_t = Y_{t-1}$，$\nabla^d = \nabla\cdots\nabla$，$\varepsilon_t$ 为白噪声，$\phi(B) = 1 - \phi_1 B\cdots - \phi_p B^p$，$\theta(B) = 1 - \theta_1 B\cdots - \theta_q B^q$，$p$ 为自回归阶数. q 为滑动平均的阶.

由于混沌与分形的发展，d 非整数的情形也已引起了人们的兴趣. 它提供了非线性建模的途径. 作为本例的特例 ARIMA（1，0，0）又记作 AR（1）为一阶自回归模型：$X_t = \phi_1 X_{t-1} + \varepsilon_t$，它与带噪声的迭代模型 $X_t = f(X_{t-1}) + \varepsilon_t$ 的区别仅在于 f 是非线性的.

在实际中大量应用的是带 Gauss 白噪声的线性的平稳时间序列模型，它的基本条件是：(1) 自回归生成函数 $A(z) = z^p - \sum_{j=1}^{p}\phi_j z^{p-j}$ 和滑动平均生成函数 $B(z) = \sum_{j=0}^{q} b_j z^{j-1}$ $(b_0 = 1)$ 的根全在单位圆内；(2) $\{\varepsilon_t\}$ 是独立同分布 $N(0, \sigma_\varepsilon^2)$.

记

$$\mu_X = \frac{\phi_0}{1 - \sum_{j=1}^{p}\phi_j}, \quad \sigma_X^2 = \frac{\sigma_\varepsilon^2}{2\pi}\int_{-\pi}^{\pi}\left|\frac{B(e^{-i\omega})}{A(e^{-i\omega})}\right|^2 d\omega,$$

则有

定理 2.1.2 若时间序列 $\{X_t, t \geqslant 0\}$ 满足 $X_0 \sim N(\mu_X, \sigma_X^2)$，则

(i) $X_t = \sum_{j=1}^{p}\phi_j X_{t-j} + \phi_0 + \sum_{j=1}^{q} b_j \varepsilon_{t-j} + \varepsilon_t$ 是平稳的；(ii) 对任何整数 t_1，…，t_k，$(X_{t_1}$，…，$X_{t_k})$ 服从多维正态分布；(iii) X_t 可表为 $X_t = \mu_X + \sum_{j=0}^{\infty}\beta_j \varepsilon_{t-j}$，其中 $\beta_0 = 1$，$\sum_{j=0}^{\infty}\beta_j^2 < \infty$；(iv) 序列有可逆性，$\varepsilon_t$ 可以有关于 X_s，$s \leqslant t$ 的线性表示.

这样的序列 $\{X_t\}$ 在时间域是完全可以由 μ_X 及自相关函数 $\rho(\tau)$ 所确定，从频率域看，则由 μ_X 及谱密度 $f(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}\left|\sum_{j=0}^{\infty}\beta_j e^{-ij\omega}\right|^2$ 所完全确定，$f(\omega)$ 还是连续的可积函数. 上述定理还揭示了线性时间序列模型与 Gauss 分布的内在联系. 读者还应了解到 ARMA 模型并非是一些零零碎碎的针对特殊情况的模型. 由于它的谱密度有 $e^{-i\omega}$ 的有理函数的形式，有理函数在连续可积函数类中足够广泛以及考虑到谱密度含有序列的大量信息，人们可以信赖 ARMA 模型的广泛性和普适性. 有了模型，统计学家要做的事是估计有限的参数 p，q 及 ϕ_i，θ_j（或 b_j）. 而这些也是面对混沌现象统计学家对非线性迭代骨架 $X_t = f(X_{t-1})$ 所要做的. 不同的对象，类似的问题！

可逆性并不总是存在的，更有甚者，1990 年在英国皇家统

计学会组织的有关机会与混沌（Chance or Chaos）的专题讨论会上，90 高龄的统计学家 Bartlelt 指出时光倒置可以建立起确定的混沌模型与随机模型之间的联系．一个有趣的例证是

例 考虑 AR（1）模型 $X_t = \frac{1}{2}X_{t-1} + \varepsilon_t, t = 1,2,\cdots$，其中 $X_0 \sim U[0,1]$，$[0,1]$ 上的均匀分布，ε_t 为两点分布取值 0 和 $\frac{1}{2}$ 的概率各为 $\frac{1}{2}$. 如果将时光倒置则这一自回归模型完全等价于一个确定性的非线性模型：Bernoulli Shift，伯努利位移. $X_{t-1} = 2X_t(\text{mod } 1), t = 0, -1, -2, \cdots$，此映射在混沌文献中讨论甚多，它是确定的但有丰富的随机性，其不变分布为 $U[0,1]$，并且关于此不变分布是遍历的.

出现这一现象是由于 AR（1）模型的映射是 1—2 的，又由于 X_0 是均匀分布使两分支有相等的概率．而在位移映射中两个不同的值可映到同一点上：比如 0.7 和 0.2 在 $2X_t(\text{mod } 1)$ 下都得到 $X_{t-1}=0.4$. 在第四章我们还将看到著名的 Logistic 映射与非线性时间序列模型在时光倒置下联系起来了．本例也体现了随机性和确定性在这特定模式中的有机统一. 回首望过去，历史轨迹，条条清晰．向前看未来，充满了未知和不确定性.

§ 2.2 布朗运动

早在 1827 年，英国植物学家 Brown 在观察悬浮在液体中的微小粒子所做的不规律运动时就发现了 Brown 运动. 它的近代研究则是由爱因斯坦和 Wiener 完成的．近年来 Brown 运动

更作为随机过程的重要基础分支而倍受重视. 它与位势理论、随机积分、统计中的不变性原理乃至本书要讨论的分形、自相似都有联系. 典型的布朗运动路径如图 2.3.

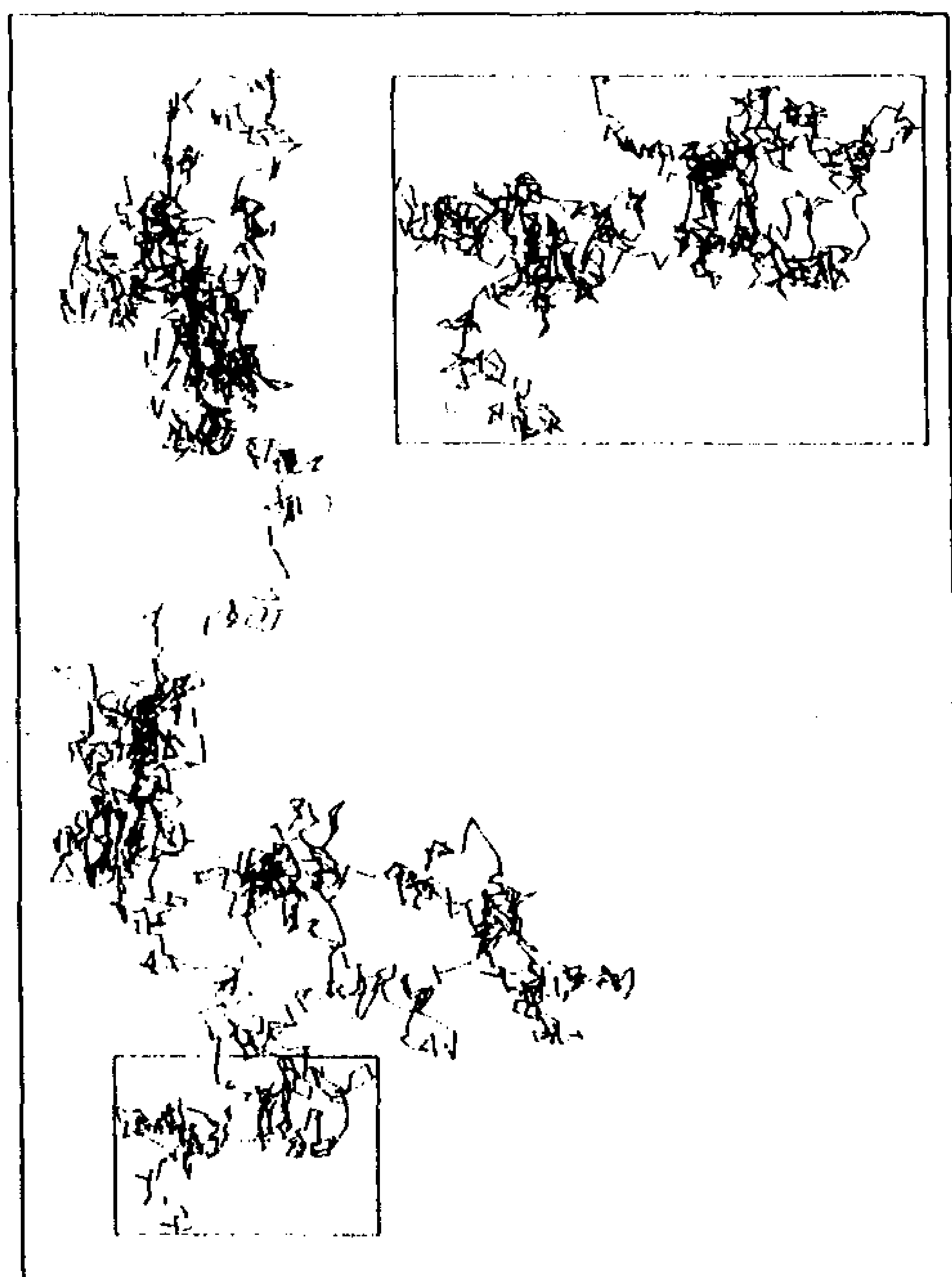


图 2.3

仍考虑直线上的随机游动模型, 假定过程在 Δt 时间单位、每次以步长 Δx 向左或向右移动的概率各为 $\frac{1}{2}$, 以 $X(t)$ 记粒子在时刻 t 所处的位置. 则

$$X(t) = \Delta x \sum_{i=1}^{N} X_i ,$$

其中 X_i 为 i.i.d 的两点分布取± 1的概率各为$\frac{1}{2}$. $N=\left[\frac{t}{\Delta t}\right]$, $[x]$ 为 x 的整数部分. 显见有 $EX(t)=0$, $\mathrm{Var}X(t)=(\Delta x)^2 N$. 若令当 $\Delta x\to 0$, $\Delta t\to 0$ 时有$(\Delta x)^2\sim c^2\Delta t$, 则有 $\mathrm{Var}X(t)\to c^2 t$. 上述随机游动模型在间隔和步长以上述量阶缩小的极限过程所得到的就是布朗运动.

定义　随机过程 $X(t)$ $(t\geqslant 0)$ 如果满足: (i)$X(0)=0$; (ii)$X(t)$ 的增量是平稳与独立的; (iii) 对 $t>0$, $X(t)$ 服从正态分布 $N(0,c^2t)$, 则称为是布朗运动, c 为 1 时又叫标准的布朗运动.

回头再看随机游动模型, 由于 $X(t)$ 是i.i.d 随机变量 X_i 的和, 故知其增量的独立性和平稳性, $X(0)=0$ 是约定, 由中心极限定理知其极限为 $N(0,c^2t)$ 随机变量. 因而是布朗运动. 记给定 $X(0)=0$, 过程在 t 单位时间后处于位置 x 的条件概率密度为 $p(x,t)$, 则有

定理 2.2.1　$p(x,t)$ 为扩散方程

$$\frac{\partial p(x,t)}{\partial t}=D\frac{\partial^2 p(x,t)}{\partial x^2}$$

的解, 其中 D 为扩散系数, 它与温度、摩擦系数及 Avogadro 常数有关, 为方便计, 常取 $D=\frac{1}{2}$.

证明　由标准 Brown 运动的定义

$$p(x,t)=\frac{1}{\sqrt{2\pi t}}\exp\left\{-\frac{1}{2t}x^2\right\},$$

直接对 $p(x,t)$ 求偏导数, 即可以验证 $p(x,t)$ 为扩散方程的解.

另一证明方法是仍从随机游动模型出发. 记

$$P(X(n)=k)=p_k(n),$$

则由全概率公式

$$\begin{aligned} & p_k(n+1) \\ = & P(X(n+1)=k \mid X(n)=k+1)P(X(n)=k+1) \\ & + P(X(n+1)=k \mid X(n)=k-1)P(X(n)=k-1) \\ = & \frac{1}{2}p_{k+1}(n)+\frac{1}{2}p_{k-1}(n), \end{aligned}$$

于是

$$p_k(n+1)-p_k(n)=\frac{1}{2}[p_{k+1}(n)-2p_k(n)+p_{k-1}(n)].$$

左边是关于时间的一阶差分，右边是关于位置的二阶差分. 取游动的时间间隔为 Δt，步长为 Δx，令 $\Delta t \to 0$ 时 $\Delta t \sim (\Delta x)^2$，并令 $k \to \infty$，$n \to \infty$时，$k\Delta x \to x$，$n\Delta t \to t$，则上述差分方程趋于偏微分方程. 由随机游动模型当 $\Delta t \sim (\Delta x)^2$ 时趋于 Brown 运动知结论为真.

这一证明思路是属于物理学家爱因斯坦的. 布朗运动有许多重要而有趣的性质，尽管粒子在瞬间的动向捉摸不定，但人们却可以从整体上把握它的运行规律.

性质 1 布朗运动 $X(t)$ 的样本路径是处处连续的，但其导数却几乎处处不存在. 连续性是显然的，不可微性的物理意义是粒子在每一瞬间都可能受到介质中分子的碰撞，其运动方向是任意的，速度为无穷. 这还可由下面的论证看出：

记 $\Delta X(t)=X(t+\Delta t)-X(t), \Delta X(t) \sim N(0,\Delta t)$，

对任何大 M，当 $\Delta t \to 0$ 时，

$$P\left\{\left|\frac{\Delta X(t)}{\Delta t}\right|\leqslant M\right\}=P\left\{\frac{\Delta X(t)}{\sqrt{\Delta t}}\leqslant M\sqrt{\Delta t}\right\}$$
$$=\Phi(M\sqrt{\Delta t})-\Phi(-M\sqrt{\Delta t})\to 0.$$
这正表明 Brown 运动在 t 的导数有限的概率趋于 0.

性质 2 记 $m(\cdot)$ 为 Lesbegue 测度，并记 $W=\frac{1}{t}m\{s\mid X(s)>0, 0\leqslant s\leqslant t\}$ 为 [0, t] 时间区间中过程处于大于 0 状态的时间比例，则有

$$P\{W\leqslant x\}=\frac{2}{\pi}\sin^{-1}\sqrt{x},\qquad 0\leqslant x\leqslant 1.$$

基于对定理 1.4.2 和随机游动与布朗运动关系的理解，这个结果是自然的.

性质 3 自相似性与分形结构. 曾在 1967 年以"英国海岸线有多长?"一文震惊学术界的科学家 Mantalbort 是分形与混沌科学的创始人之一. 他曾说过："布朗运动的轨迹是大自然本身描绘的曲线". 实际上 $d\geqslant 2$ 维空间的布朗运动其值域集合的分形维数均为 2. 当人们改变对布朗运动观察采样的时间间隔，所得的路径都有相同程度的复杂性. 其形状在统计意义上是自相似的，并有分形的特性. 这种自相似性可表述为.

定义 一个随机过程满足 $Y(ct)\overset{d}{=}C^{H}Y(t)$，则称 $Y(t)$ 为有自相似系数 H 的过程.

布朗运动 $X(t)\sim N(0,t)$，故 $X(ct)\sim N(0,ct)$. 而 $\sqrt{c}X(t)\sim N(0,(\sqrt{c})^2t)=N(0,ct)$，故有 $X(t)$ 的自相似系

数为$\frac{1}{2}$.

布朗运动还有一些变种也都十分有用. 若 $X(t)$ 为标准布朗运动,则 $Y(t)=|X(t)|$ 称作反射布朗运动;$Y(t)=\exp\{X(t)\}$ 则为几何布朗运动. 后者在数理经济学中可用来模拟股票市场的价格波动. 记 X_i 为均值 0、方差 1 的独立随机变量, 记

$$X_n(t)=\frac{1}{\sqrt{n}}\sum_{i=1}^{[nt]}X_i+\frac{nt-[nt]}{\sqrt{n}}X_{[nt]+1},$$

则 $X_n(t)$ 在一定的函数空间中以分布弱收敛到布朗运动 $X(t)$. 这是著名弱不变性原理的内涵. 它将统计学中的极限定理的证明从零散的单个序列的处理跃进为成批的整体处理. 又因为上述结果不依赖 X_i 的原始公共分布,便可以通过对特殊相对简单的独立和序列极限定理的证明而获取复杂分布的相应结果.

定义 $X(t)$ 为[0,1]上的布朗运动,对 $0\leqslant t\leqslant 1$,记 $B(t)=X(t)-tX(1)$,则 $B(t)$ 称为是布朗桥过程.

容易算出 $EB(t)=0$,及 $\mathrm{Cov}(B(s),B(t))=s\wedge t-st$,$s\wedge t$ 表两者中小的数. 因 $B(1)=B(0)=0$,在[0,1]两端 $B(t)$ 是固定的而得名. 它在过程收敛中扮演重要角色. 以经验分布函数为例, 设 F 为 [0, 1] 区间上的均匀分布, X_1, X_2, …, X_n 为从 F 中抽取的 i.i.d 样本, $F_n(s)=\frac{1}{n}\sum_{i=1}^{n}1_{[x_i\leqslant s]}$ 为经验分布函数, 则由中心极限定理, 过程 $B_n(s)=\sqrt{n}\,F_n(s)-s$ 依分布收敛到 $N(0,s(1-s))$. 从 $F_n(s)$ 的分子服从二项分布 $B(n,s)$ 可知 $E[F_n(s)]=\frac{1}{n}\sum_i F(s)=F(s)=s$ 及 $\mathrm{Var}[F_n(s)]=\frac{1}{n}s(1-s)$, 故 $B_n(s)$ 收敛到 $B(s)$. 据此可以证明统计中十分重要的

Glivenko 定理

$\sup\limits_{0\leqslant s\leqslant 1}|F_n(s)-F(s)|$ 以概率趋于 0.

布朗运动的方差为 c^2t，或者说与时间间隔 $|t|$ 成比例. 它假定不同时间区间上过程的行为是统计独立的. 在现实数据处理中人们把微小的相互关联近似地看作不相关，并就长期而言，相信遥远过去的因素对当今的事件影响甚微. 但不少科学家也发现了一种长程相依. 在长期水文、气象资料研究领域也已发现这种关联. 于是可用分数布朗运动来描述. 这类过程是 Kolmogrov 在 1940 年提出的，Mandelbort 等人在 1968 用它研究了水文资料.

定义 若随机过程 $X_H(t)$ 满足：(i) $X_H(0)=0$；(ii) $EX_H(t)=0$，$\mathrm{Var}X_H(t)=|t|^{2H}$；(iii) 增量 $X_H(t+s)-X_H(t)$ 与 t 无关（有平稳性）服从正态分布，则称之为分数布朗运动.

不难算出，$X_H(t)$ 之自相关函数 $\rho(\tau)=1+(2H-1)\tau^{2H-2}$，$H$ 就是自相似系数，$0<H<1$，当 $H>\frac{1}{2}$，$\sum\limits_k\rho(k)=\infty$，表现出过程可能以很长的周期在重复. 由于人们对自然现象的观察只能在有限的时间中进行，采集的资料也是有限的，便很难断言是否存在比观察期间长得多的周期. 如何区分混沌和长周期现象也是混沌科学中的难点之一. 通过倍周期分叉走向混沌的道路是漫长的，而人类走出混沌的道路将更加漫长.

§2.3 收敛性与遍历性

在§1.2中，读者已接触到大数定律和中心极限定理. 这些都是概率论中各种收敛性的实例. 与通常分析中的极限不同的是这些收敛性需用概率的语言加以描述.

设 $\{X_n, n\geqslant 1\}$，X 是概率空间 Ω 上的随机变量序列，则有

定义 当 $P\{\lim\limits_{n\to\infty} X_n = X\} = 1$，称 X_n 以概率1收敛于 X. 也称作几乎处处或几乎必然收敛，常称为强收敛；若对任给的 $\varepsilon>0$ 有

$$\lim_{n\to\infty} P(|X_n - X| \geqslant \varepsilon) = 0,$$

则称 X_n 依概率收敛到 X.

定义 若对 $r\geqslant 1$，$X_n - X$ 的 r 阶绝对矩的极限为0，

$$\lim_{n\to\infty} E|X_n - X|^r = 0,$$

则称 X_n r 阶平均收敛于 X. 当 $r=1$ 为 L_1 模收敛，$r=2$ 为均方收敛.

定义 若记 F_n，F 分别为 X_n 及 X 的分布函数，且对 F 的每个连续点 x 有

$$\lim_{n\to\infty} F_n(x) = F(x),$$

则称 X_n 依分布收敛到 X，或称 X_n 的分布 F_n 弱收敛于 X 的分布 F.

分布弱收敛的一个有用的等价条件是对任何有界连续函数

都有

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} g(x)dF_n(x) = \int_{-\infty}^{\infty} g(x)dF(x).$$

上述这些收敛性之间的关系为当 $r>s$ 时 r 阶平均收敛蕴含着 s 阶平均收敛. 且 L_1 模收敛蕴含依概率收敛. 而以概率为 1 收敛蕴含着依概率收敛,依概率收敛蕴含着依分布弱收敛. 而 r 阶平均收敛与以概率为 1 收敛则互不包含. 从蕴含关系中可约略体会强何以为强，弱何以为弱. 由于收敛类型的不同，在遍历性的讨论中 (定理 2.3) 也将有各种不同的遍历性.

遍历性，又称各态历经性是研究保持测度不变的可测变换的渐近性状的数学分支,它源起于统计力学的一个基本假设:一个真实的平衡的物理系统在某时刻的状态与相应的稳定系综(ensemble),即一切可能的微观系统的全体,在相空间的点有相同的概率，遍历性理论与动力系统理论、概率论、信息论、数论各个科学分支都有密切的联系. 用一句远非准确的简单话语来形容、所谓遍历性就是“相平均=时间平均”这在定理 2.1.1 中已遇到过了. 在这个意义上，人们也看到了时间空间的转换与有机统一. 遍历论的早期成果有 Birkhoff 1931 证明的遍历性定理和 Von Neumann 的平均遍历性定理，这些都是划时代的开创性工作,到了本世纪 60 年代,微分动力系统的遍历性质的研究受到重视. 数学家证明了整个流型上有双曲结构的系统是遍历的. 之后, S. Smale, R. Bowen 和 D. Ruelle 将此结果推广到了所谓公理 A 系统. S. Smale 和 D. Ruelle 两位学者都是混沌学研究历史上响哨哨的人物. 而 Bowen 却不幸于 1978 年 7 月 30 日英年早逝、去世时年仅 31 岁,实为遍历性研究领域的重大损失.

样本空间 Ω 中可以赋以概率的那些事件的集合数学上抽

象为σ一代数，记为$\mathscr{F}$，$(\Omega,\ \mathscr{F},\ P)$. 这个三元组就是抽象的概率空间，它是我们讨论遍历性的出发点. 若$X(\omega)$为一随机变量，则集合$\{\omega|X(\omega)\leqslant x\}$,$x$为实数,就是$\mathscr{F}$中的集合. $P\{\omega|X(\omega)\leqslant x\}$记为$F(x)$，就是熟知的分布函数.

定义 变换T：$\Omega\to\Omega$称为是可测的，如对每一个$A\in\mathscr{F}$有$T^{-1}A=\{\omega|T(\omega)\in A\}\in\mathscr{F}$.

定义 若T为一可测变换，且对任何$A\in\mathscr{F}$，$P(T^{-1}A)=P(A)$，则称T为保测变换.

从定义可知所谓保测变换是指T不仅可测，而且使$\mathscr{F}$中的集合在变换之后保持测度P(或者我们要讨论的概率测度)不变. 保测变换与平稳序列有着很自然的联系. 如果T为一保测变换，定义随机变量$X_k(\omega)=X_1(T^{k-1}\omega)$，则随机序列$X_1,X_2,\cdots$是平稳序列；反过来，对$(\Omega,\ \mathscr{F},\ P)$上的每一个平稳序列$\{X_k,k\geqslant1\}$人们都能构造出一个新的$(\overline{\Omega},\overline{\mathscr{F}},\overline{P})$及随机变量序列$\overline{X}_k(\overline{\omega})$，及一个保测变换$S$，使$\{\overline{X}_1(\overline{\omega}),\overline{X}_1(S\overline{\omega}),\cdots,\overline{X}_1(S^{k-1}\overline{\omega})\cdots\}$与$\{X_k,k\geqslant1\}$有相同的分布!其中的$S$就是著名的后移变换,$S:R^\infty\to R^\infty$满足$S(x_1,x_2,\cdots)=(x_2,x_3,\cdots)$.

例 令$\Omega=[0,1)$,$\mathscr{F}=[0,1)$上Borel集全体,P为通常的Lebesgue测度,对任何$\alpha\in[0,1)$,$T(x)=x+\alpha\pmod 1$及$T(x)=2x\pmod 1$均为保测交换. 用统计学的语言讲若$X\sim U(0,1)$,则变量$X+\alpha\pmod 1$,$\alpha\in[0,1)$及$2X\pmod 1$均仍服从均匀分布$U[0,1)$. 但$T(x)=\frac{1}{2}x$就不是保测变换.

早在 1912 年 Poincare 就证明了有关保测变换的重要常返性定理.

定理 2.3.1 (Poincare) 令 $(\Omega, \mathscr{F}, P)$ 为一概率空间，T 为一保测变换，$A \in \mathscr{F}$，则对几乎所有的 A 中的 ω，使 $T^n\omega \in A$ 的 n 有无穷多个.

证明 记 $B = \{\omega \in A \mid T^n\omega \notin A, n \geqslant 1\}$，则有 $B \cap T^{-n}B = \varnothing$，于是 $T^{-m}B \cap T^{-(m+n)}B = T^{-m}(B \cap T^{-n}B) = \varnothing$. 故 $\sum_{n=0}^{\infty} P(B) = \sum_{n=0}^{\infty} P(T^{-n}B) \leqslant P(\Omega) = 1$. 因此 $P(B) = 0$. □

下面转入遍历性的定义和遍历性定理的叙述.

定义 T 为保测变换，集合 $A \in \mathscr{F}$ 若满足 $T^{-1}A = A$，则称为是不变的（关于 T）. 可以证明不变集的全体构成一个 σ-代数记作 $\mathscr{I}$. 若对所有 $\omega \in \Omega$ 有 $X(\omega) = X(T\omega)$，则称 X 是不变的随机变量.

定义 $(\Omega, \mathscr{F}, P)$ 上的保测变换 T 若满足对每一个不变集 A 有 $P(A) = 0$ 或 1，则称 T 是遍历.

例 当且仅当 α 为有理数 $T(x) = x + \alpha \pmod 1$ 是遍历的.

定义 一个保测变换 T 称为是有混合性的，若对任意的 $A, B \in \mathscr{F}$ 都有

$$\lim_{n \to \infty} P(A \cap T^{-n}B) = P(A)P(B).$$

它表明事件 A 与 $T^{-n}B$ 当 $n\to\infty$ 时渐近独立，这反映了长期系统演化后的混乱性状.

在动力系统的讨论中，Ω 就是系统的状态空间，ω 为初始时的状态、$T^n\omega$ 为时刻 n 时的状态. $T^n\omega$ 是由运动方程所确定的从状态空间到其自身的变换，当 $T^{n+m}\omega = T^n(T^m\omega)$ 时系统是保守的. $X(\omega)$ 为状态 ω 的可观察函数. 在物理学的实际测量中观察时间刻度相对于分子间互相影响的精细自然时间刻度要长得多. 所以实际中得到的只是系统演化过程中不同状态观察的平均 $\frac{1}{n}\sum_{k=1}^{n}X(T^{k-1}\omega)$. 物理学家 Gibbs 猜想 $T^n\omega$ 会在状态空间中游荡，而在每一小邻域中 $T^n\omega$ 会趋于一个极限分布 P. 而关于时间的平均将应等于相空间的平均 $E_p(X(\omega))$. 这就是 Birkhoff 在 1931 年证明的遍历性定理的大意.

定理 2.3.2 T 为概率空间 $(\Omega, \mathscr{F}, P)$ 上的保测变换，对满足 $E|X(\omega)| < \infty$ 的 $X(\omega)$，几乎处处有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}X(T^k\omega) = E(X|\mathscr{I}).$$

其中 $E(X|\mathscr{I})$ 代表给定不变集 $\mathscr{I}$ 后 X 的条件期望. 当 T 为 $(\Omega, \mathscr{F}, P)$ 上的保测变换且是遍历的，则对满足 $E|X| < \infty$ 的 X 几乎处处有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}X(T^k\omega) = EX.$$

对于随机序列 X_n，当它是平稳的，我们有

定义 记 B 为 R^∞ 中的 Borel 集，一个集 $A\in\mathscr{F}$ 当可表为 $A=\{\omega|\ (X_n, X_{n+1}, \cdots)\in B\}$ 时，称为关于平稳序列 $\{X_k\}$

不变的. 不变集的全体记为$\mathscr{I}_X$. 一个平稳序列当其任何不变集的测度均为 0 或 1 时，称为是遍历的.

相应于定理 2.3.2 的有

定理 2.3.3 $X_1, X_2, \cdots$为一个平稳序列，且$E|X_1|<\infty$，则几乎处处和依均方收敛有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_k = E(X_1|\mathscr{I}_X).$$

若$X=\{X_1,X_2,\cdots\}$是遍历的，则有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_k = EX_1.$$

它的一个重要的推论是若$X_1, X_2, \cdots$为独立同分布的随机变量，则序列是平稳的和遍历的. 于是几乎处处有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_k = EX_1.$$

这就是著名的 Kolmogrov 强大数律的充分性部分. 可以证明对 Gauss 平稳序列当$\lim\limits_{\tau\to\infty}\rho(\tau)=0$（渐近不相关）时是遍历的. 这个条件是容易验证的. 下面给出定理 2.3.2 的证明大意. 这一简化证明是 A. Garsia 1965 年给出的.

证明 首先不妨假$E(x|\mathscr{I})=0$,不然可考虑序列$Y=X-E(X|\mathscr{I})$,记$L_n=\sum\limits_{k=0}^{n-1}X(T^k\omega)$. 为证定理只需证几乎处处有

$$0\geqslant\overline{\lim_{n\to\infty}}\frac{L_n}{n}\geqslant\underline{\lim_{n\to\infty}}\frac{L_n}{n}\geqslant 0.$$

记$\overline{\lim\limits_{n\to\infty}}\dfrac{L_n}{n}$为$Z$,则由$Z(\omega)=Z(T\omega)$知$Z$是不变的,于是集合$A_\epsilon$

$=\{\omega | Z(\omega) > \varepsilon\}$ 也是不变的. 只要能证明对任何 $\varepsilon > 0$, $P(A_\varepsilon) = 0$ 则必有 $P\{Z \leqslant 0\} = 1$. 类似可以证明 $\lim\limits_{n\to\infty} \frac{L_n}{n} \geqslant 0$ 几乎处处成立,定理遂可得证. 为证 $P(A_\varepsilon) = 0$,记 1_{A_ε} 为 A_ε 的示性函数,于是由 $E(X | \mathscr{I}) = 0$ 知

$$
\begin{aligned}
& -\varepsilon P(A_\varepsilon) = 0 - \varepsilon P(A_\varepsilon) \\
& = E[E(X | \mathscr{I})]1_{A_\varepsilon} - \varepsilon P(A_\varepsilon) \\
& = E[X(\omega)1_{A_\varepsilon}] - \varepsilon P(A_\varepsilon) \\
& = E[(X(\omega) - \varepsilon)1_{A_\varepsilon}].
\end{aligned}
$$

为方便计,记上面括号中的随机变量 $(X(\omega) - \varepsilon)1_{A_\varepsilon}$ 为 Y,并记 $S_k(\omega) = Y(\omega) + \cdots + Y(T^{k-1}\omega)$; $M_k(\omega) = \max(0, S_1, \cdots, S_k)$. 由 $S_1 = Y$ 及 $M_n \geqslant 0$ 知 $Y \geqslant S_1(\omega) - M_n(T\omega)$, 及 $Y(\omega) + M_n(T\omega) \geqslant Y(\omega) + S_k(T\omega) = S_{k+1}(\omega)$. 所以 $Y(\omega) \geqslant \max(S_1(\omega), \cdots, S_n(\omega)) - M_n(T\omega)$. 再注意到在 $\{M_n > 0\}$ 上 $M_n = \max(S_1(\omega), \cdots, S_n(\omega))\}$, 因而

$$
\begin{aligned}
& E[Y(\omega)1_{[M_n>0]}] \\
& \geqslant E\{[\max(S_1(\omega), \cdots, S_n(\omega))] - M_n(T\omega)\} \\
& \geqslant E\{(M_n(\omega) - M_n(T\omega))1_{[M_n(\omega)>0]}\} \\
& \geqslant E\{M_n(\omega) - M_n(T\omega)\} = 0.
\end{aligned}
$$

最后期望等于 0 利用了 T 的保测性. 最后再验证 $[M_n > 0]$ 作为集合序列当 $n \to \infty$ 时是单调上升到 A_ε 的,由实分析中的控制收敛定理. 通过期望号取极限即可证明 $E((X - \varepsilon)1_{A_\varepsilon}) \geqslant 0$. □

保测变换是相对三元组 $(\Omega, \mathscr{F}, P)$ 而言的,若给定了可测空间 $(\Omega, \mathscr{F})$ 及变换 T, 则使 T 成为保测变换的 P 就叫做不变概率测度. 对某一个 T, P 不是唯一的. 如果不变概率测度不能分解成两个不相等的不变概率测度的线性组合, $P = \alpha P_1$

$+\beta P_2, P_1 \neq P_2$、$\alpha+\beta=1$，则称之为不可分解的，一个不可分解的不变概率测度就是遍历的概率测度. 若同一个遍历的保测变换 T 有两个不变测度 P_1、P_2，则要么 $P_1=P_2$，要么 P_1 与 P_2 是正交的（所谓正交是指互为奇异：测度 μ 关于 ν 奇异是指在 ν 测度非零的集合 S_ν 上 μ 测度为 0，$S_\nu \cap S_\mu=\varnothing$，常记为 $\mu \perp \nu$）. T 的不变概率测度的全体构成一个凸集，此凸集的极点正是关于 T 的遍历测度. 对于遍历概率测度，若 φ 为连续函数，则几乎处处有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\varphi(X(T^k\omega)) = E_P\varphi(X).$$

取 $\varphi\equiv1$，又回到定理 2.3.2.

例 把 $[0,1)$ 中的点写成二进制 $x=0.x_1x_2\cdots$，并使其表示无歧义. 考虑变换 $T(x)=2x(\mathrm{mod}\ 1)$，这时 $T(0.x_1x_2\cdots)=0.x_2x_3\cdots$. 取 p，$0<p<1$，定义 X_i 为两点分布，取值 0，1 的概率为 p 及 $1-p$ 且相互独立. 则 P_p 是关于位移变换 T 的不变测度且是遍历的，变换 T 实际上就是一个动力系统. 可见对不同的 p 就有不可数多个不同的遍历测度，且互为奇异.

讨论遍历性与本书主题混沌的关系何在？在统计科学中遍历性的研究使人们能把握随机性，而确定性模型出现混沌现象并呈随机特性也提供了把握随机性的有用模式；在混沌的数学定义中有一条拓扑传递性，其表现出来的是一种混合性，而本节定义的混合变换蕴含遍历性，两者有些共同之处；混沌学研究中的若干数量特征的确定也有赖遍历性定理.

定理 2.3.4 每一个混合的保测变换都是遍历的.

证明 取 $A\in\mathscr{F}$，$B\in\mathscr{I}$，则对 $n\geq1$，$T^{-n}B=B$. 所以对

$n \geqslant 1$ 有

$$P(A \cap T^{-n}B) = P(A \cap B).$$

由混合性定义

$$P(A \cap B) = \lim_{n \to \infty} P(A \cap T^{-n}B) = P(A)P(B).$$

特别取 $A=B$，则有 $P(B) = P^2(B)$，于是 $P(B) = 0$ 或 1. 所以任何不变集 B 的概率测度为 0 或 1. 故 T 是遍历.

行文至此，我们为混沌学的讨论作了必要的铺垫. 这些蜻蜓点水式的铺垫在一定程度上简略介绍了概率论与随机过程若干基本概念. 后两章将进入主题.

第三章　混　　沌

气似质具而未相离谓之混沌．

《易学象数论·乾凿度》

跨越世纪的两位人物曾有过这样的争论：Laplace 在 1776 写道“如果想象一位大智大慧者能在某一时刻把握宇宙间实体的全部关系，那么他就能讲出所有这些实体在过去与未来的相对位置、运动以及广泛的相互影响．”而伟大的法国数学家 Poincare 辨驳说：“事实并非如此．初始条件微小的差别可能对最终的现象产生非常大的影响，前面的小误差会酿成以后的大偏差，预言是不可能的！”Poincare 起码在关于初始条件细小差异会给动力系统未来演化带来巨大影响的思想方面是深邃而先知先觉的．这番话是 1903 年讲的．七八十年之后在混沌科学中大放异彩，成了轰动科学界的新思维．所谓混沌就字面本意而言，是杂乱无章，混混沌沌，极端紊乱无序．它是系统复杂和行为不可预测的同义语．混沌现象的令人吃惊之处是一个完全确定的简单模型也会产生有如随机过程的行为特性．如果你有

幸发现这些相对简单的模式对不确定的未来的预测竟有可能实现.

有关混沌的数学研究的历史可以追溯到 1890 年左右，Poincare 有关太阳系稳定性的研究. 他曾问道：行星是会大致继续沿着目前的轨道不停地运行呢？还是会向黑暗深邃的太空逃逸亦或堕向太阳而毁灭呢？他本人虽未能对此作出回答，却创立了动力系统几何学的研究方法. 前苏联数学家 Kolmogrov 也对动力系统的不规则性作了深入的研究，有名的 KAM 定理就是以他的名字命名的. 该定理主要论述 Hamilton 系统方程组解的稳定性，证明了几乎可积系统有填满不变环的拟周期解存在. 这在微分方程和天体力学两方面都是有关太阳系稳定性理论的重大突破. KAM 三人中的另两位是 Arnold 和 Moser. 到了 60 年代，美国数学家 Stephen Smale 研究了微分动力系统结构的不稳定性. 他试图对动力系统行为性状的典型类型进行分类，混沌即是其中之一. 著名的马蹄映射也以他的名字命名. 1963 年，Lorenz 在关于气象预报的简单微分方程的研究中发现系统轨线依赖精确的初始条件的混沌现象. 结果以后被人们广为引用. 1975 年李天岩与 York 在美国数学月刊上发表了“周期 3 隐含着混沌”一文，首次把混沌 (Chaos) 作为一个数学定义提出来. 同时欧洲数学家 David Ruelle 和 Floris Takens 对混沌现象作了大量的研究. 到 80 年代，有关混沌与分形的著作如雨后春笋. 1987 年 Gleick 的一本《混沌学》成为全球的畅销书，使混沌在科学与知识界尽人皆知. 自然界的分形随处可见，云彩、山川、河流、地形地貌都因分形的引入而更加多姿多采. 混沌与分形在物理、生物、化学、气象、水文、地震、医学乃至中华气功各个领域的研究蓬蓬勃勃，势不可挡. 有兴趣的读者不妨去看看 Gleick 那本脍炙人口的科普佳作. 而本书则主要从统计学者

的眼光来看混沌与混沌科学．本章的四节大体可用四句话来概括：混沌是对初值敏感；周期 3 蕴含着混沌；混沌映射有正的 Lyapunov 指数，其奇异吸引子常有分形结构；混沌是信息源，体现不确定性的拓扑熵大于 0.

§3.1 数学家看混沌

本节从非线性动力系统的复杂变化来引入混沌学的有关重要概念．记 $N=\{0,1,2,\cdots\}$，考虑时间序列 $\{x_n, n\in N\}$，若假定 x_0 为初始条件，且 $x_{n+1}=f(x_n)$，其中 f：$D\to D$ 为映射．D 一般为一个距离空间的紧子集，$f(\cdot)$ 为非线性函数．在迭代过程中可能出现混沌现象．

$x_{n+1}=f(x_n)$ 也称作是一个离散的动力体系，它可以嵌入在叫做流的动力系统 $\varphi(t,x)$ 之中（φ 满足 $\varphi(0,x)=x$；$\varphi(s+t,x)=\varphi(s,\varphi(t,x))$．它也是微分动力体系 $\frac{dx(t)}{dt}=F(x(t))$．$t\in R$ 的离散形式．x_n 也可看作 x_0 经过 n 次迭代后的值，即 $x_n=f^n(x_0)$，其中 $f^n(x)=f(\underbrace{f\cdots f}_{n}(x))$．有关函数迭代的研究有了近百年的历史，$f^n(x)$ 的解析形式一般很难获得．只有在计算机科学高度发展的今天，数学家才有可能看到当 n 很大时 $f^n(x)$ 所展现出来的奇妙情景．已有人在倡导计算机上的“实验数学”．

1．对初值的敏感

为简单计，先看直线上的连续可微函数的迭代．

定义 集合 $\{x, f(x), f^2(x), \cdots\}$ 称为在迭代映射 f 下 x 的向前轨道.

轨道中序列的性质是迭代研究中所关心的问题.

定义 若 x_0 满足方程 $f(x) = x$,则称 x_0 是 f 的不动点.若有 $f^n(x_0) = x_0$，且对 $1 \leqslant k < n$, $f^k(x_0) \neq x_0$，则 x_0 为 f 的 n-周期点. 由 n-周期点的所有迭代组成的集合称为是 n-周期轨.

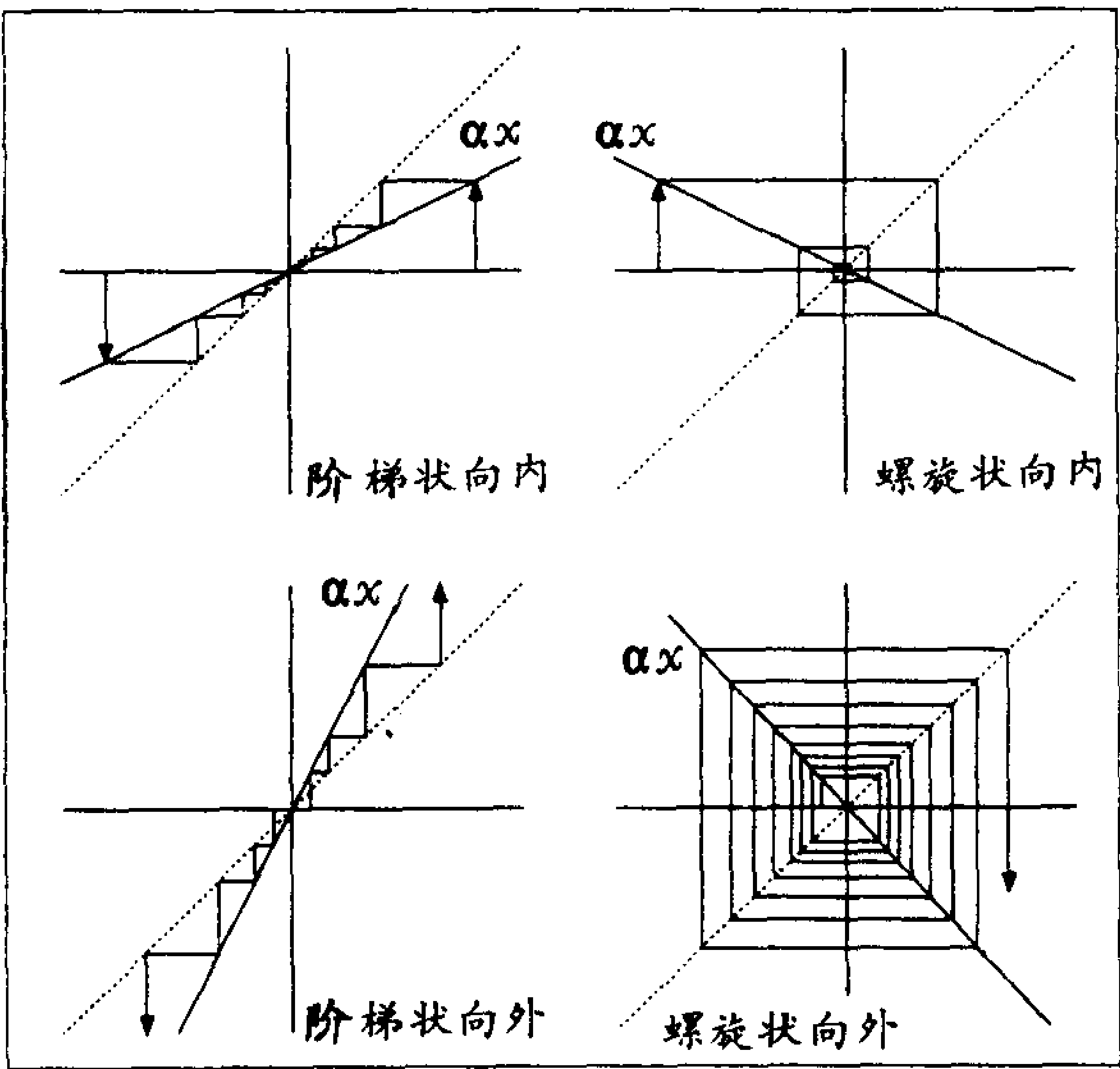


图 3.1

比如,$\{x_0, f(x_0), \cdots, f^{n-1}(x_0)\}$ 就是 f 的 n-周期轨. 1-周期轨就是不动点,2-周期轨只有两点 $(x, f(x))$，把它们看作平面上的点,这时关于直线 $y=x$ 对称的两点 $(x, f(x))$ 及 $(f(x), x)$

均在函数 f 的图形上. 利用图形可以对动力系统在不动点周围的变化情形得到直观的印象. 这种方法叫图分析.

例 观察迭代 $f(x)=\alpha x$，0 为不动点. 如图 3. 1 所示，当 $\alpha\in(0,1)$ 时迭代阶梯式的趋向稳定的不动点 0，当 $\alpha\in(-1,0)$ 时迭代向内螺旋式地趋向稳定不动点 0. 当 $\alpha>1$ 或 $\alpha<-1$，迭代分别以阶梯式或螺旋式向外的方式远离不稳定的不动点而去. $\alpha=1$ 时，所有点都是不动点，$\alpha=-1$，$(x,-x)$为 2-周期轨.

定义 对不动点 p，若 $|f'(p)|\neq 1$，则点 p 称为是双曲点，当 $|f'(p)|>1$，p 为排斥不动点；当 $|f'(p)|<1$，p 为吸引不动点，$|f'(p)|=1$ 时 p 为驻点. 对 n-周期点可类似定义它的双曲性. 吸引周期点也叫做陷井.

定理 3.1.1 若 p 为吸引不动点，则存在一个含 p 的开区间 U，使得当 $x\in U$ 时有 $\lim\limits_{n\to\infty}f^n(x)=p$.

证明 由于 f 连续可微，存在 $\varepsilon>0$，使当 $x\in[p-\varepsilon,p+\varepsilon]$ 时，$|f'(x)|<A<1$. 由中值定理，

$$\begin{aligned}|f(x)-p|&=|f(x)-f(p)|\\&\leqslant A|x-p|<|x-p|\leqslant\varepsilon.\end{aligned}$$

所以 $f(x)\subset[p-\varepsilon,p+\varepsilon]$，且比 x 离 p 更近. 类似有

$$|f^n(x)-p|\leqslant A^n|x-p|.$$

当 $n\to\infty$ 有 $\lim\limits_{n\to\infty}f^n(x)=p$.

考虑带参数 μ 的迭代函数族 $f_\mu(x)=\mu x(1-x)$，(在不引起混淆时也可直接简记为 f). 这一族映射被生物学家称作是

Logistic 映射. 当 μ 变化时迭代的性状会发生有趣的变化. 现观察 f_μ: $D\to D$, $D=[0,1]$, 不动点为 0 及 $p_\mu = 1-\frac{1}{\mu}$. 当 $\mu<1$, 0 为吸引不动点, 且当 $0<x\leqslant 1$ 有 $\lim\limits_{n\to\infty} f^n(x)=0$. 当 $\mu=1$, 0 为唯一的不动点且 $f'(0)=1$, 故为非双曲情形; 当 $\mu>1$, 0 为排斥吸引点. 当 $1<\mu<3$ 时, p_μ 为吸引不动点对 x 属于 $(0,1)$ 有 $\lim\limits_{n\to\infty} f^n(x)=p_\mu$. 当 $\mu>3$, 情形将变得复杂: 分叉、倍周期和混沌相继出现, 此乃后话. 为方便计算, 令 $\mu=4$. 在

$$x_{k+1}=f(x_k)=4x_k(1-x_k)$$

的迭代过程中一旦 x_k 的计算发生小误差 ε_k, 则 x_{k+1} 的误差有多大呢? 这容易算:

$$\varepsilon_{k+1}=f(x_k+\varepsilon_k)-f(x_k)=4\varepsilon_k(1-2x_k)-4\varepsilon_k^2,$$

于是

$$\varepsilon_{k+1}/\varepsilon_k=4(1-2x_k)-4\varepsilon_k,$$

其主项由 $f_4(x)$ 在 x_k 的导数 $4(1-2x_k)$ 所决定. 因此 $f'(x_k)$ 将代表一次迭代后误差的放大倍数, 它一般依赖于 x_k 的位置, n 次迭代后的误差增长为

$$\frac{\varepsilon_n}{\varepsilon_0}=\frac{\varepsilon_n}{\varepsilon_{n-1}}\cdot\frac{\varepsilon_{n-1}}{\varepsilon_{n-2}}\cdots\frac{\varepsilon_1}{\varepsilon_0}.$$

取其绝对值几何平均的对数为 $\frac{1}{n}\sum\limits_{k=1}^{n}\log\left|\frac{\varepsilon_k}{\varepsilon_{k-1}}\right|$. 上式极限当 $n\to\infty$存在时可有

定义 迭代 f 在 x_0 的 Lyapunov 指数为

$$\lambda(x_0)=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\log|f'(x_{k-1})|.$$

若 x_0 为 f 的 m-周期点，则 $\lambda(x_0)=\frac{1}{m}\sum_{k=1}^{m}\log|f'(x_{k-1})|$. 特别若 x_0 为不动点，$\lambda(x_0)=\log|f'(x_0)|$. $x_0=0$ 为 f_4 之不动点，故 $\lambda(0)=\log4$（亦有 $\lambda(1)=\log4$），实际上更准确地说，对所有其轨道最终能达到 0 的点 x，均有 $\lambda(x)=\ln4$；而对 $[0,1]$ 中其他几乎所有 x（其 Lebesgue 测度为 1），则有 $\lambda(x)=\log2\doteq 0.693$. 因对 $[0,1]$ 中绝大部分 x，每迭代一次近似于误差加倍 ($e^{0.7}=2$). 而 n 次迭代后的误差增长为

$$dx_n=e^{\lambda n}dx_0.$$

上式所表达的含义是系统演化对初始条件的变化十分敏感. 失之毫厘，差之千里.

关于映射 $f(x)=4x(1-x)$ 的遍历测度为反正弦分布，其密度为 $p(x)=\frac{1}{\pi}\frac{1}{\sqrt{x(1-x)}}$. 可以通过计算机实验来检验这一点. 对初始值 $x_0=0.20005$，将 4000 次迭代的结果依位置和频数描出 U 型直方图如图 3.2a）所示. 若在 $[0.10005, 0.30005]$ 中取 4000 个初值，并将其在第 2000 次迭代后在相空间的结果依频数点成另一个 U 型直方图如图 3.2b）所示. 两者的惊人相似是由遍历性定理所保证的. 于是在初始值 x_0 的任一小邻域中都有点 y，当迭代次数足够多之后，其向前轨道与 x_0 的向前轨道严重偏离. 这种对初值的敏感使确定性迭代的结果变得难以预料（unpredictable). 遍历性使这一迭代映射表现出某种“随机性”. 遍历性定理还保证几乎处处有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\log|f'(f^k(x))|=E[\log|f'|].$$

上述极限将不依赖于 x，记为 λ. 它正是映射 $f(x)=4x(1-x)$ 所对应的 Lyapunov 指数. 它代表了随着迭代次数的增加，迭代在时间方向上的平均拉伸程度. Lyapunov 指数 $\lambda>0$ 所表示的

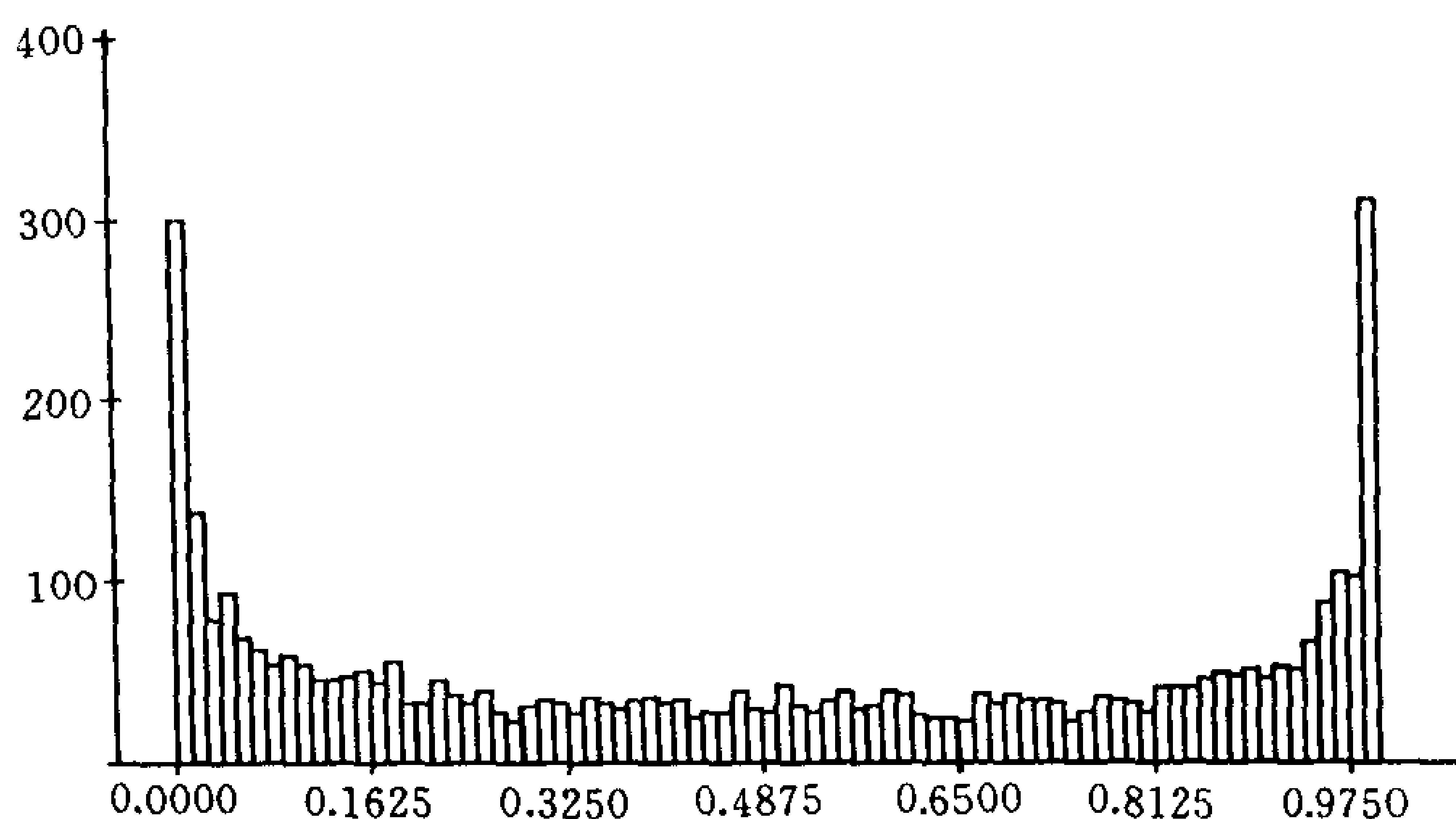


图 3.2 (a)

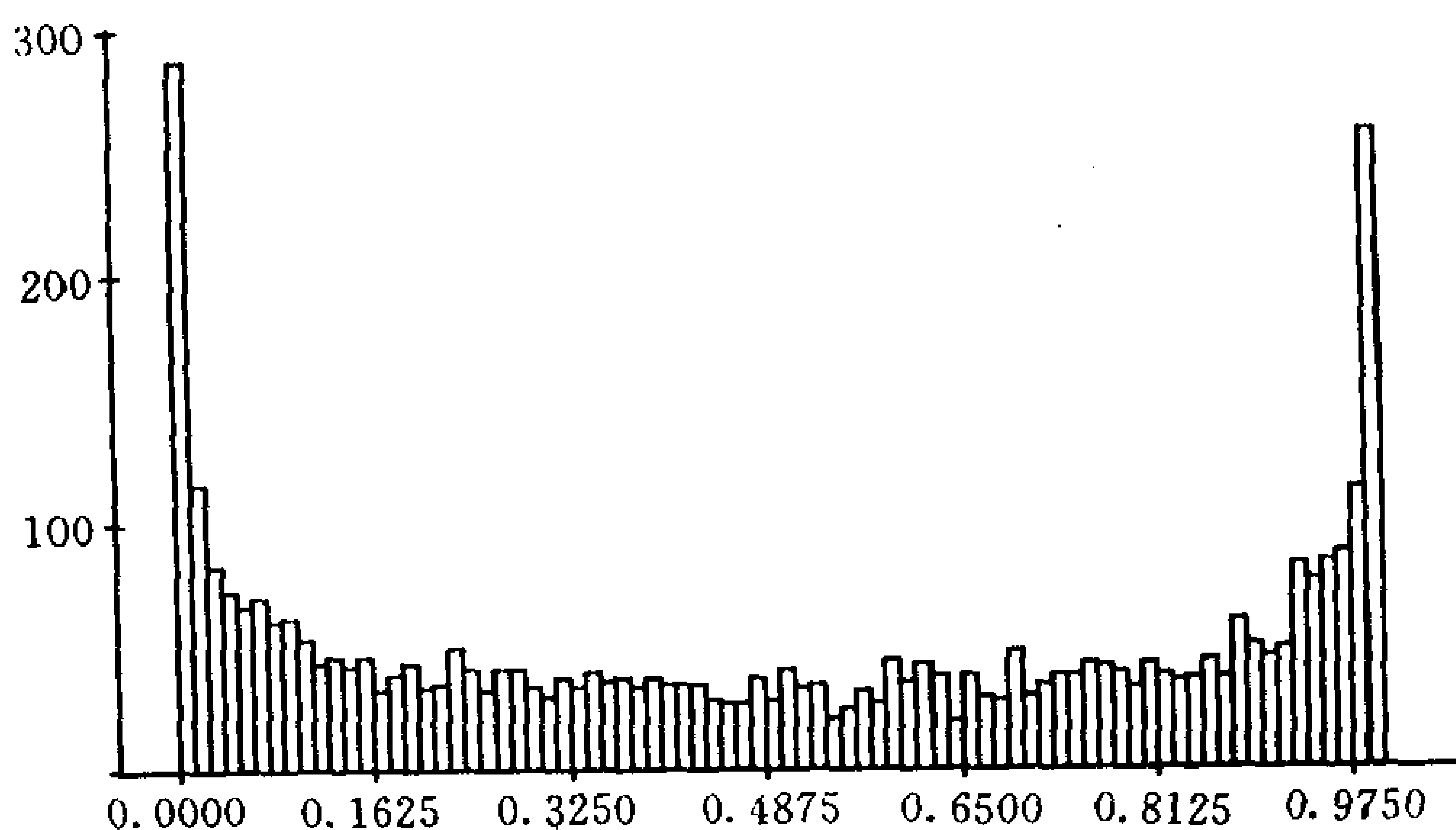


图 3.2 (b)

对初始值敏感是混沌的重要特征之一. 容易算出

$$E[\log|f'|]=\int_0^1\log|4-8x|\frac{1}{\pi\sqrt{x(1-x)}}dx$$

$$=\log 2>0.$$

当 $\lambda<0$ 时迭代所产生的效应是沿着向前轨道压缩，收敛到不动

点也是可能的情形之一.

2. 混合性（mixing）

换一个角度来看[0,1]区间上的 Logistic 映射. 发现有如揉面团那样的混合性：对任何两个[0,1]中的开区间 I 及 J，都可以找出 I 中的点为初始值经多次迭代后最终会进入 J. 对于一个给定的作为初始点的小区间 I 和一个小的目标区间 J，从 I 中均匀取点经过迭代考查它们是否最终进入目标区间 J，一旦进入就算该初值消亡. n 次迭代后仍未能进入 J 的称作成活者，记成活者的数目为 $S(n)$. $S(n)$ 以指数速度衰减，即 $S(n) \propto e^{-\frac{n}{\tau}}$. 计算机实验中取 $I=[0.2, 0.2+10^{-11}]$，$J=[0.68, 0.69]$，τ 为 140 左右. τ 称为轨道的平均寿命时间，它的值依赖于 I 的位置. 依据不变分布反正弦律我们可以对 τ 的大小进行理论估计：每次迭代后落入 J 的概率为

$$p=\int_{0.68}^{0.69}\frac{1}{\pi\sqrt{x(1-x)}}dx \doteq 0.0068527.$$

于是为使 τ 次迭代后成活率为原来的 $\frac{1}{e}$，则应有 $(1-p)^{\tau}=\frac{1}{e}$. 因此 τ 的估计值为 $\hat{\tau}=145$，与实验结果接近. 更有甚者：任取一个小区间，并观察其逐次迭代，在此过程中区间的平均大小将增加并在有限次迭代后填满整个[0,1]区间. 例如直接计算就知道 $f^5([0.1,0.2])=[0,1]$. 一般而言，起始区间长度减半，所需平均迭代的次数加1，这只不过是 $e^{\lambda}=2$ 的另一种表述方式. 要注意迭代 f 的周期点及其原象都只能在有限的周期轨内游荡，并不能在揉面过程中混合到所有任给的其他小区间去. 从保持概率不变出发把目标区间视为随机变量的取值，访问这些区间的频数可近似变量落在目标区间的概率，则可导出迭代

的不变分布所应满足的方程. 由 $f(x)=f(1-x)$ 知 f 是关于 $x=\frac{1}{2}$ 对称的. 记不变分布的密度为 $\nu(x)$，随机变量 $Y=4X(1-X)$，则

$$P\{Y\in J\}=\nu(y)dy.$$

而当 $X\in I_1$ 及与 I_1 关于 $\frac{1}{2}$ 对称的部分 I_2 都能使 $Y\in J$，于是对 $x_1\in I_1$ 及 $x_2=(1-x_1)\in I_2$ 有

$$\nu(y)dy=\nu(x_1)dx_1+\nu(x_2)dx_2.$$

由 $dy=f'(x)dx$ 代入左边可得

$$\nu(y)=\frac{\nu(x)}{|f'(x)|}+\frac{\nu(1-x)}{|f'(1-x)|}.$$

这是有名的 Frobenius-Perron 方程，其解为

$$\nu(x)=\frac{1}{\pi\sqrt{x(1-x)}}.$$

在实际中，不变分布可以从一初始概率密度出发，利用 Frobenius-Perron 算子 $G(\nu_k)=\nu_{k+1}$ 求出 ν_k. 一般而言，ν_k 收敛到不变分布密度 ν. 其中

$$G(\nu)(y)=\sum_{x\in f^{-1}(y)}\frac{\nu(x)}{|f'(x)|},y\in[0,1].$$

上述混合性在数学上描述为

定义 映射 f: $D\to D$ 如满足对 D 中的任一对开集 A 及 B 都存在 $k>0$ 使 $f^k(A)\cap B\neq\varnothing$，则称 f 是拓扑传递的.

拓扑传递性所表现出来的是确定性映射 f 的随机行为性状. 但这种紊乱绝非一片混沌，形无序而实有序. 这体现在 f 的周期点集在 $[0,1]$ 中稠密. 为使有关讨论方便，需要引进符号

动力系统和拓扑共轭的概念，而他们本身也是动力系统领域的重要论题.

3. 周期点稠密

讨论迭代 f 的周期点稠密乃至混沌都可以借助于对帐篷变换、位移映射的研究而获得. 这类变换有如面包师揉面、拉伸、卷叠不断揉合. 混沌也被人们称为是面包师的杰作.

定义 T：[0, 1] → [0, 1],

$$T(x)=\begin{cases}2x, & x\leqslant\frac{1}{2},\\ -2x+2, & x>\frac{1}{2}.\end{cases}$$

如图 3.3 所示. 因其形如帐篷而得名. T 对 [0,1] 的作用酷似面包师揉面. 在不断的拉伸、卷叠；再拉伸、再卷叠下经多次迭代混合得以实现.

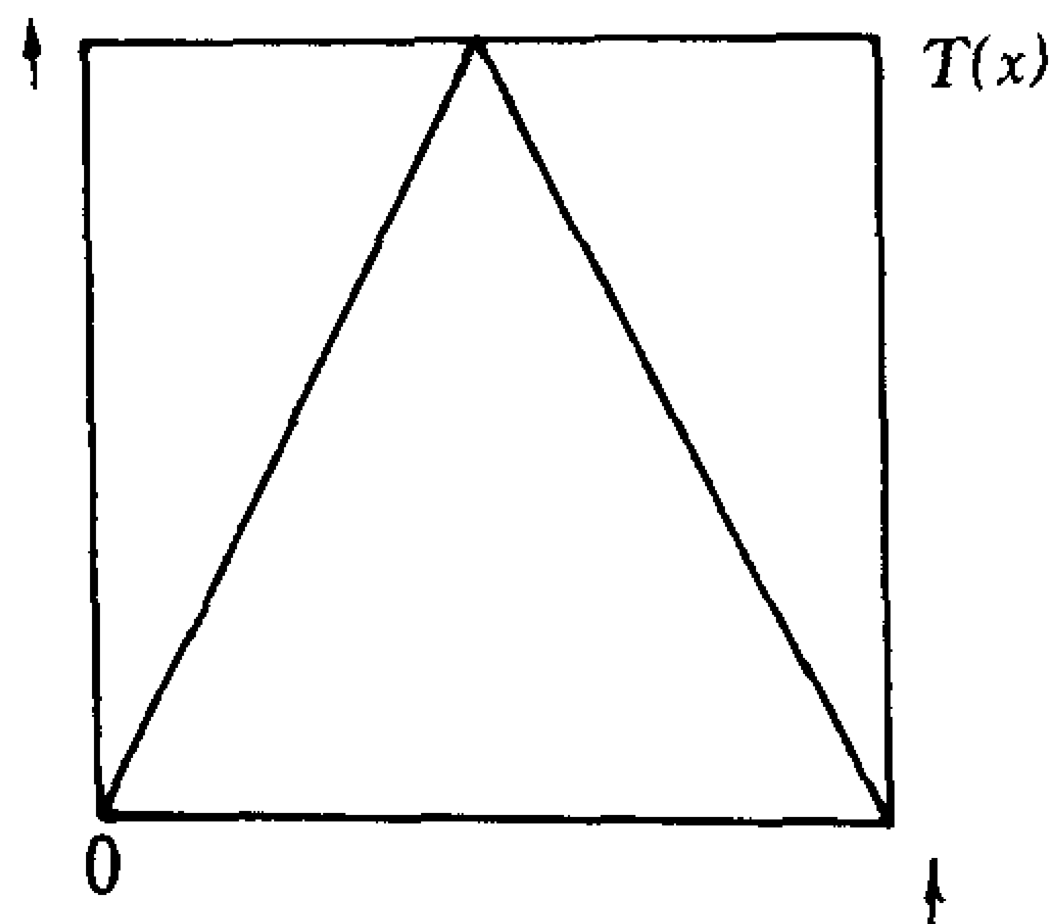


图 3. 3

定义 S:[0,1]→[0,1],

$$S(x)=\begin{cases}2x, & x<\frac{1}{2},\\ 2x-1, & x\geqslant\frac{1}{2},\end{cases}$$

称为锯齿变换.

S 的作用是延展、拉伸、切割再粘贴. S 又可表为 $S(x)=2x\ (\mathrm{mod}\ 1)$，如第二章中所看到的它等价于当 x 为二进制小数表示时的位移映射.

定理 3.1.2 $TS^{n-1}=T^n$.

证明　当 $n=1$，2，可直接验证．再由归纳法

$$\begin{aligned}TS^n &= TS^{n-1}\cdot S = T^nS\\ &= T^{n-1}\cdot TS = T^{n-1}\cdot T^2 = T^{n+1}.\end{aligned}\quad\square$$

借助这一定理可容易直接求出 T^k：

$$T^k(x_0) = TS^{k-1}(x_0) = T(y),$$

其中
$$y = S^{k-1}(x_0) = 2^{k-1}x_0 \pmod 1.$$

定理 3.1.3　变换 S 对初值敏感、有混合性，且周期点稠密．

证明　对任给 $\varepsilon>0$ 及 $\Delta<\frac{1}{2}$，取整数 k 使 $2^{-k}<\varepsilon$．取初始值 x_0 的二进制小数表示使它仅精确到小数之后 k 位 $0.a_1a_2\cdots a_k$，并从 a_{k+1}起由扔钱币的正反来确定其值为 1 或 0．取 y_0 仅在第 $k+1$ 位上与 x_0 不同，则 $|y_0-x_0|\leqslant\frac{1}{2^k}<\varepsilon$（差别微小）．但 $S^k(x_0)=0.a_{k+1}a_{k+2}\cdots$；$S^k(y_0)=0.a^*_{k+1}a_{k+1}\cdots$，$|S^k(x_0)-S^k(y_0)|=\frac{1}{2}>\Delta$．这表明对初值的敏感．再看混合性：对任取两个小区间 I 及 J，小区间的长度 $|I|>\frac{1}{2^{n-1}}$，取其中点表为 $0.a_1a_2\cdots a_n\cdots$，$y=0.b_1b_2b_3\cdots$ 为 J 中一点．则 $x_0=0.a_1a_2\cdots a_nb_1b_2b_3\cdots$必属于 I，但 $S^n(x_0)=y$ 已属于 J 了．记 ω_0 为二进循环小数 $0.\dot{a}_1\dot{a}_2\cdots\dot{a}_k$，其中 $a_1\cdots a_k$ 为循环节．$S^k(\omega_0)=\omega_0$，所以 ω_0 为 S 的 k-周期点．对任给的 $x_0=0.a_1a_2a_3\cdots$及任给的 $\varepsilon>0$，存在 k 使 $2^{-k}<\varepsilon$ 及 $\omega_0=0.\dot{a}_1\cdots\dot{a}_k$，使 $|\omega_0-x_0|\leqslant\varepsilon$．所以周期点是稠密的，遍布在任意选取的点 x_0 的邻域内．这和无理数可用连分数逼近的道理是一样的．

定理 3.1.4 帐篷变换 T 也对初值敏感，有混合性且周期点稠密.

证明 若 $x=0.a_1a_2a_3\cdots$，则当 $x<\frac{1}{2}$ 时有

$$T(x)=S(x)=0.a_2a_3\cdots;$$

当 $x\geqslant\frac{1}{2}$, $S(x)=2x-1$,

$$\begin{aligned}T(x)&=2-2x=1-(2x-1)\\&=1-S(x)=1-0.a_2a_3a_4\cdots\\&=0.a_2^*a_3^*a_4^*\cdots,\end{aligned}$$

其中 $a^*=1_{[a=0]}$，与 a 的取值相异.

若 ω 为 S 的 n-周期点则 $T(\omega)$ 为 T 的 n-周期点，此因

$$T^n(T(\omega))=T(S^n(\omega))=T(\omega).$$

所以，为证 T 之周期点稠密性，对任意 $x_0=0.a_1\cdots a_n\cdots$，取 $\omega=0.\dot{0}\dot{a}_1\cdots\dot{a}_n<\frac{1}{2}$ 为 S 的 $n+1$-周期点，而 $T(\omega)=0.\dot{a}_1\cdots\dot{a}_n\dot{0}$ 即为所求的 $n+1$-周期点 ω_0，它与 x_0 之距离可小于任给的 ε. 读者不妨自行证明其他两条性质.

在研究函数迭代时 f^n 的表达式一般很难写出. 但若能有 A、B 两拓扑空间（比如 $[0,1]$ 区间）间的连续一一映射 h 使 $h\circ f=g\circ h$（$\circ$ 表复合运算），则 $f=h^{-1}\circ g\circ h$. 并且 $f^n=h^{-1}\circ g^n\circ h$. 借助于 g^n 就有可能把 f^n 表出. 方程 $h\circ f=g\circ h$ 又称为 Schröder 方程，已有一百多年的历史了. 把这种技巧抽象为一种概念就是

定义 设 A、B 是两个拓扑空间，f: $A\to A$ 和 g: $B\to B$ 分别是 A 与 B 上的连续自映射. 如果存在 A 到 B 的同胚 h: $A\to$

B 使得

$$h \circ f = g \circ h,$$

则称 f 与 g 是拓扑共轭的. 同胚 h 叫 f 到 g 的一个拓扑共轭.

拓扑共轭是 f 与 g 之间的一种等价关系. 有关同胚及拓扑共轭的深入讨论可参看有关动力系统的著作. 对我们来说一个重要的事实是 f 与 g 有相同的动力学性质，若 g 有稠密的周期点，则 f 也同然. 依 Banks 等人 1992 年在美国数学月刊上发表的结果还可知：若 g 是混沌的，f 关于 g 自共轭则 f 也是混沌的（自共轭的概念放宽了自映射连续的条件）. 而对 h 的要求是一一到上连续且逆函数连续.

例 取 $f(x) = \frac{4}{5}(x-1)$，$h(x) = x + 4$，

$$h^{-1}(x) = x - 4, \quad g(x) = \frac{4}{5}x,$$

则

$$f(x) = h^{-1} \circ g \circ h(x).$$

由 $g^n(x) = \left(\frac{4}{5}\right)^n x$ 容易求出

$$f^n(x) = h^{-1}\left(\left(\frac{4}{5}\right)^n h(x)\right) = \left(\frac{4}{5}\right)^n (x+4) - 4.$$

例 取 $h(x) = \sin^2\left(\frac{\pi x}{2}\right)$，则 Logistic 映射 $4x(1-x)$ 关于帐篷变换 T 是拓扑共轭的.

定理 3.1.5 Logistic 映射 $f(x) = 4x(1-x)$ 是混合的，且周期点是稠密的.

证明 取 $y \in [0, 1]$，记 $x = h^{-1}(y)$，由定理 3.1.4，T

的周期点在 $[0,1]$ 中稠密，故存在 T 的 p_k-周期点序列 x_k 使 $\lim\limits_{k\to\infty}x_k=x$. 取 $y_k=h(x_k)$，则亦有 $\lim\limits_{k\to\infty}y_k=y$，且由

$$f^{p_k}(y_k)=f^{p_k}(h(x_k))=h(T^{p_k}(x_k))=h(x_k)=y_k,$$

知 y_k 为 f 的 p_k-周期点. 取 I，J 为 $[0,1]$ 中两开区间由 h 连续知 $h^{-1}(I)$ 及 $h^{-1}(J)$ 仍为开集. 由 T 是混合的，对 $x\in h^{-1}(I)$ 存在 k 使 $T^k(x)\in h^{-1}(J)$. 取 $y=h(x)\in I$，由 $f^k(y)=f^k(h(x))=h(T^k(x))\in J$ 便知 f 亦是混合. 仍借助 J. Banks 的结果：X 为实直线上的任一子集，f：$X\to X$ 为具有混合性的周期点稠密的连续变换，则 f 亦有对初始条件的敏感性. 即可知 Logistic 映射也如我们已看到的那样对初始值敏感.

至此可以综合这三条为

定义 （Devaney）设 D 是一个度量空间，映射 f：$D\to D$ 如满足下列三条件，则称 f 在 D 上混沌：

(i) 存在 $\delta>0$，使对任给 $\varepsilon>0$ 和任意 $x\in D$，有 $y\in D$ 及 $n\geqslant 0$，使 $\|x-y\|<\varepsilon$，$\|f^n(x)-f^n(y)\|>\delta$，其中 $\|\cdot\|$ 为 D 中之距离；

(ii) f 有拓扑传递性；

(iii) f 的周期点在 D 中的稠密.

依此定义，定理 3.1.5 告诉我们映射 $4x(1-x)$ 在 $[0,1]$ 上是混沌的. 本节讨论混沌的方法有其普遍意义. 由 0 和 1 构成的无穷序列的全体组成的集合

$$\Sigma_2=\{(a_0a_1a_2\cdots)\mid a_i=0\text{ 或 }1\}$$

称为 $\{0,1\}$ 上的序列空间. 而 Σ_2 上动力体系的研究称为是符号

动力体系. Σ_2 上的 $(a_0a_1a_2\cdots)$ 与 $(b_0b_1b_2\cdots)$ 之距离定义为 $\sum_{k=0}^{\infty}\frac{|a_k-b_k|}{2^k}$. 位移映射 $\sigma:\Sigma_2\to\Sigma_2$ 是指

$$\sigma((a_0a_1a_2\cdots))=a_1a_2a_3\cdots$$

它在上述距离下在 Σ_2 上是连续的. σ 是 Σ_2 上的混沌，为证某个映射 f 在 D 上混沌只需证明 f 在 D 上和某个序列空间上的位移映射拓扑共轭，这便是本节所用方法的本质. 但发现同胚 h 并非轻而易举.

§3.2　倍周期、分叉及 Feigenbaum 常数

关于映射族 $f_\mu(x)=\mu x(1-x)$ 的故事尚未完结，当 $0<\mu<3$ 迭代的动力学性质较为简单；而当 $\mu=4$ 时 f 已呈混沌性状，那么当 $\mu\in[3,4)$ 区间时是怎样从简单走向复杂的呢？本节将讨论这条通过倍周期和分叉走向混沌的道路.

当 $\mu>3$，0 及 $P_\mu=1-\frac{1}{\mu}$ 均为排斥不动点，在 $[0,1]$ 中除了 P_μ 的可列个原象（存在 n 使 $f^n(x)=P_\mu$ 的 x 称作原象）之外，其他点都不会被吸引到 P_μ 去. 0 的原象为 0 和 1；1 则没有原象. 别的点经过迭代跑到哪儿去了？考虑 f 的 2-正周期点，它是四次方程 $f^2(x)=x$，也即

$$\mu[\mu x(1-x)][1-\mu x(1-x)]-x=0$$

的解. 除解 0 及 P_μ 外另有两解

$$u=[\mu+1+\sqrt{\mu^2-2\mu-3}]/2\mu,$$

$$l=[\mu+1-\sqrt{\mu^2-2\mu-3}]/2\mu.$$

且有 $f(u)=l$，及 $f(l)=u$. 当 $3<\mu<1+\sqrt{6}\doteq3.44949$ 时，

$$|f^{2'}(u)| = |f^{2'}(l)| < 1.$$

由定理 3.1.1 知 (u,l) 组成由两点构成的吸引子. 而在 $\mu=3$ 处, $|f'(P_3)| = 1$, $u=l=P_3$, 吸引点 P_3 一分为二发生第一次分叉 (bifurcation). 当 $\mu=1+\sqrt{6}$ 时 $|f^{2'}(u)| = 1$, u 又一分为二; l 也同然. 于是二点吸引子分叉为 4 点的吸引子. 周期加倍的过程可以不断继续. 20 世纪 70 年代中期物理学家 M. Feigenbaum 对此映射族进行了细致的研究. 记 $b_1=3$, $b_2=1+\sqrt{6}$, …, b_n 为使 f_{b_n} 首次有 2^n-周期点的分叉点. 记 $\lim\limits_{n\to\infty} b_n = b_\infty \doteq 3.56944$, 当 $b>b_\infty$, 情况更加复杂, 继续出现 2^kP (P 为素数 k 为整数) 周期点. 最后当 b 接近 3.83 时周期 3 出现了! 一旦有了 3-周期点, 即 $f^3(x) = x$ 有非平凡解, 则对所有 n, $f^n(x) = x$ 也有非平凡解. 这就是 Li 和 York1975 年发表在美国数学月刊上的著名论文"周期 3 蕴含着混沌"所揭示的真理. 确而言之有

定理 3.2.1 (Li-York) 若 $f(x)$ 为线段 I 中的连续自映射, 若 f 有 3-周期点, 则对一切正整数 n, f 有 n-周期点.

证明 微积分中的著名的中介值定理是连续映射的重要特点, 依赖它可以证明: 当 $[a, b] \subset I$, $f([a,b]) \supset I$ 时必有 $x_0 \in [a, b]$ 使 $f(x_0) = x_0$. 又若 $[a, b]$, $[c, d]$ 为 I 的两个子区间, $f([a,b]) \supset [c,d]$, 则$\exists$ 一个闭区间 $J \subset [a,b]$, 使 $f(J) = [c,d]$ 且 J 的任何真子区间无此特性. 更进一步, 若有 I 的一串闭子区间 J_0, J_1, …, J_{n-1}使 $f(J_k) \supset J_{k+1}$, $k=0, 1, 2, \cdots, n-1$, 则存在 $x_0 \in J_0$ 使 $f^n(x_0) = x_0$, 且对 $k=0, 1, 2, \cdots, n-1$, $f^k(x_0) \in J_k$. 现设 $\{x_1, x_2, x_3\}$ 为 f 的 3-周期轨, 则 $f(x_2) = x_1$ 或者 $f(x_2) = x_3$, 不妨设 $f(x_2) = x_1$, 于是 $f(x_1)$

$= x_3$ 及 $f(x_3) = x_2$. 记 $K_0 = [x_1, x_2]$，$K_1 = [x_2, x_3]$，则当取 $J_0 = J_1 = \cdots = J_{n-2} = K_0, J_{n-1} = K_1$. 不难验证 $f(J_k) \supset J_{k+1}$ 成立，故有 $x_0 \in J_0$，使 $f^n(x_0) = x_0$，且 $f^k(x_0) \in K_0$，$f^{n-1}(x_0) \in K_1$. 于是 x_0 为周期点，为证 x_0 是 n-周期点，只需证 $\{x_0, f(x_0), \cdots, f^{n-1}(x_0)\}$ 两两不同确实组成了 n-周期轨. 现反证这一点：若 x_0 的周期小于 n，则 $f^{n-1}(x_0)$ 必与某一个 $f^k(x_0)$ $(k=0,\cdots, n-2)$ 相同（约定 $f^0(x_0) = x_0$）. 所以 $f^{n-1}(x_0) \in K_0$，于是 $f^{n-1}(x_0) \in K_0 \cap K_1 = \{x_2\}$，故又有 $x_0 = f^n(x_0) = f(f^{n-1}(x_0)) = f(x_2) = x_3 \notin K_0$. 此为矛盾！ □

倍周期分叉的情况可以从图 3.4 中清楚地看出，图中横坐标是参数 μ 从 0 到 4 变化，纵坐标是不动点在 [0，1] 中的位置. 很像树干和树枝、最后是密密麻麻的复杂的树叶.

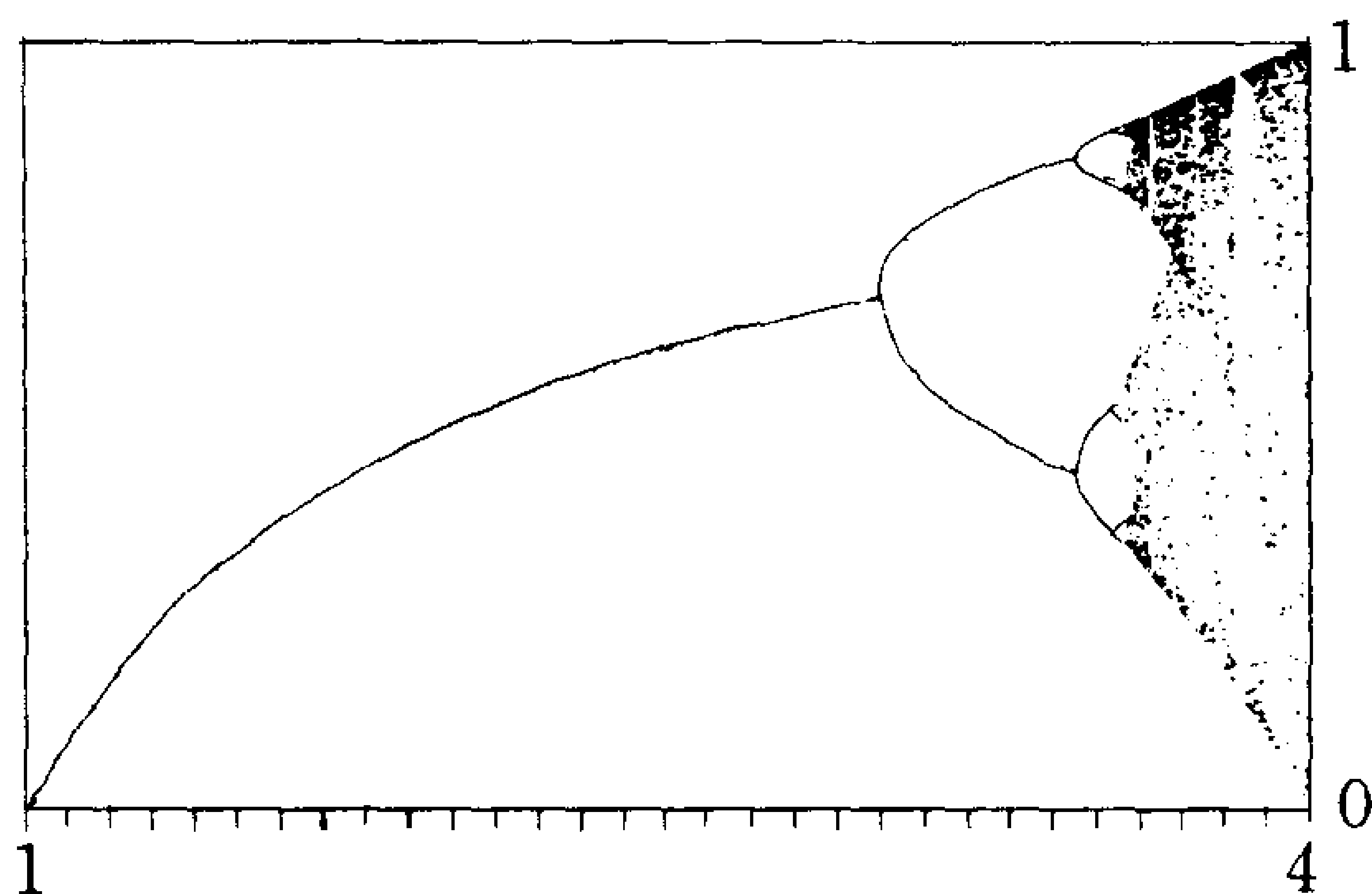


图 3.4

Feigenbaum 还发现分叉点序列之差衰减得很快，其比值的极限存在记为 δ，即有

$$\lim_{k\to\infty}\frac{b_{k+1}-b_k}{b_{k+2}-b_{k+1}}=\delta,$$

$\delta \doteq 4.6692016091029\cdots$. 时间是 1975 年 10 月. δ 也因此被命名为 Feigenbaum 常数. 有趣的是 δ 不仅适用于 Logistic 迭代，而且对满足一定条件的迭代函数类是普适的. 所需条件是：

a）f：$[0, 1] \to R$ 的光滑函数（连续、可微）；

b）f 单峰，在 f 的最大值处的二阶导数 $f''(M) \neq 0$；

c）对所有 $x \in [0,1]$，f 的 Schwarzian 导数

$$S_f(x)=\frac{f'''(x)}{f'(x)}-\frac{3}{2}\left(\frac{f''(x)}{f'(x)}\right)^2<0.$$

因其普适性有人甚至将它与常数 π、e 相提并论，足见对 δ 的重视程度.

分叉点序列的计算并不容易，Feigenbaum 引入了超吸引序列 S_k，$k=1, 2, \cdots$，其中 S_k 为关于 μ 的方程

$$f_\mu^{2^{k-1}}\left(\frac{1}{2}\right)=\frac{1}{2}$$

之解. 而有趣的是

$$\lim_{n\to\infty}\frac{S_n-S_{n-1}}{S_{n-1}-S_{n-2}}=\delta.$$

量 $S_\infty=\lim\limits_{n\to\infty}S_n=3.56994$，又称作 Feigenbaum 点，它是混沌的入口处. 在 S_∞ 左边的是倍周期树干与树枝；在它右边则是混沌所管辖的树叶. Feigenbaum 的图呈现一种惊人的自相似性，对固定的 $\mu>S_\infty$ 在纵截线上形成有如 Cantor 集的结构. 人们甚至算出此类 Cantor 集的分维维数约为 0.538，所有这些细致的研究仍要借助于符号动力体系.

对 $\mu\in(S_\infty,4]$，我们从图 3.5（a）（b）中可以清楚地发现在 $\mu=4$ 有一段含盖整个 $[0, 1]$ 区间的带子，之后随 μ 减小而变狭直至 m_1 处分成两段. 在 $(m_1, 4]$ 之间从右到左排着窗口，

越来越细；最右边是 3-周期窗、5、7 周期窗依次变狭…．从 m_1 向左 μ 继续减小到 m_2 分成四段、中间的周期窗的周期分别为 6，10，14…．在 [0，1] 区间随 μ 的减小而成倍的分段过程中又产生序列 m_k，$\lim_{k\to\infty} m_k = m_\infty = S_\infty$，且仍有（依 Gropmann 和

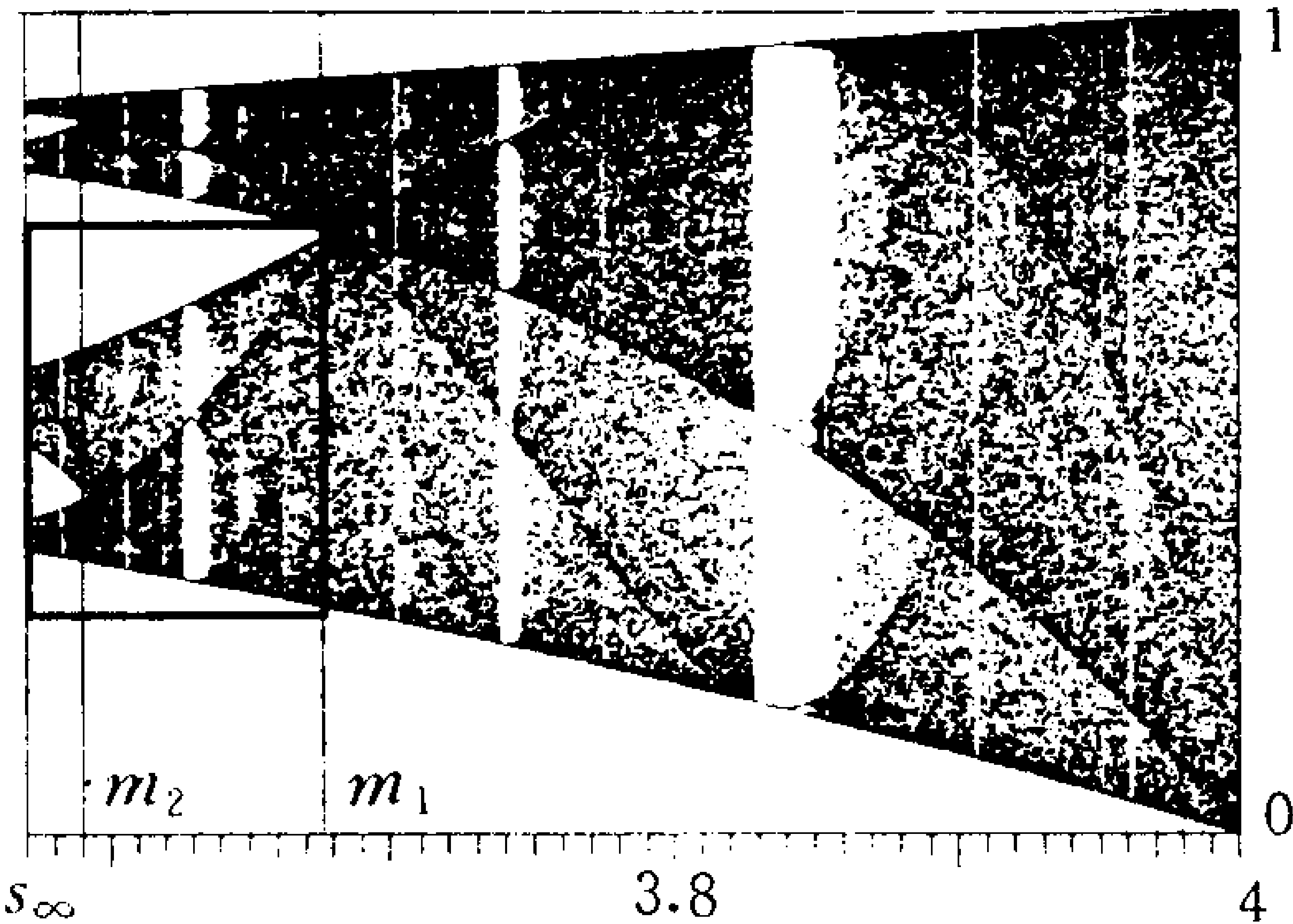


图 3.5（a）

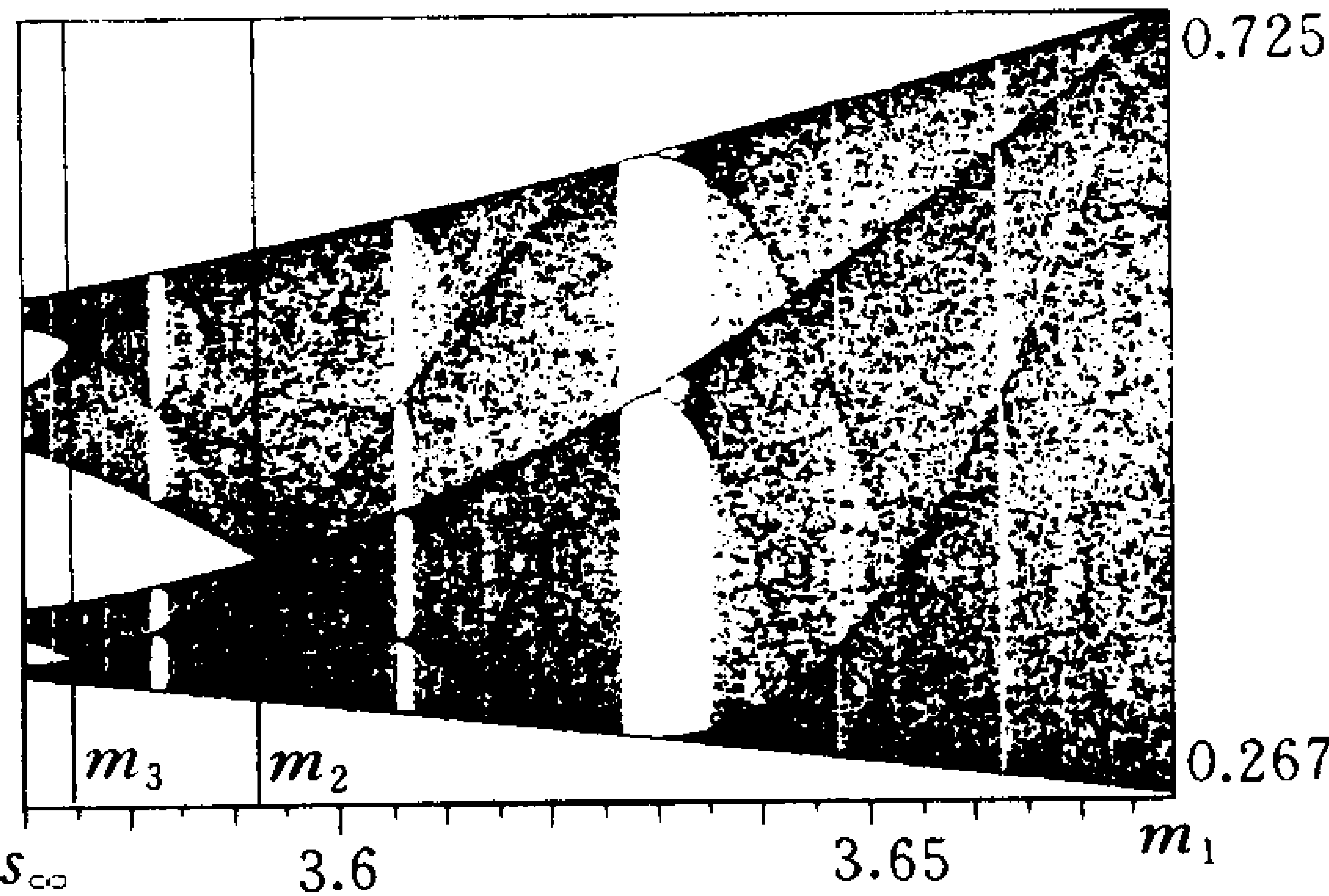


图 3.5（b）

Thomae 的实验结果)

$$\lim_{k\to\infty}\frac{m_{k+1}-m_k}{m_{k+2}-m_{k+1}}=\delta!$$

当人们为这些图形的美妙而惊讶时，却发现早在 Li-York 轰动科学界的著名论文发表前 10 年，前苏联数学家 A. N. Sarkovski 就发现过一个更加深刻而惊人的 Sarkovski 定理. 他引入◁序，将全体 自然数重新排序为

$$3\lhd 5\lhd 7\cdots\lhd 2n+1\lhd 2n+3\cdots\lhd,$$

$$2\cdot 3\lhd 2\cdot 5\lhd 2\cdot 7\cdots\lhd 2\cdot(2n+1)\lhd 2\cdot(2n+3)\cdots\lhd,$$

$$\cdots\cdots$$

$$2^k\cdot 3\lhd 2^k\cdot 5\lhd 2^k\cdot 7\cdots\lhd 2^k\cdot(2n+1)\lhd 2^k\cdot(2n+3)\cdots\lhd,$$

$$\cdots\cdots$$

$$2^l\lhd 2^{l-1}\lhd\cdots\lhd 2^4\lhd 2^3\lhd 2^2\lhd 2\lhd 1.$$

请读者欣赏这一排序方式中各排之间的相似性.

定理 3.2.2 设 $f(x)$ 为线段 I 上的连续自映射 f 有 m-周期点，则当 $m\lhd n$ 时 f 有 n-周期点.

取 $m=3$ 的特例就是 Li-York 定理. 真是山外有山，天外有天. 此定理的简化证明可参看张景中、熊金城的有趣新作:《函数迭代与一维动力系统》. 四川教育出版社，1992 年出版.

§3.3 奇异吸引子、分形及其维数

一维迭代映射已经如此丰富，高维微分动力系统如耗散动力系统所呈现的画面将更加迷人. 一个微分动力系统可由下式

描述：

$$\frac{dX(t)}{dt} = F(X(t)), \quad t > 0.$$

$X(0)$ 为初始条件，$X(t)$ 为系动在时刻 t 的状态 $X(\cdot)$ 是多维的，F 也是多维的，其值域为状态空间. 方程的特解在状态空间中描出一个轨道，当 $\sum_{i=1}^{k} \frac{\partial F_i}{\partial x_i} < 0$ 时，耗散系统的轨道就被吸引到称为奇异吸引子的集合上. 这些轨道也呈现对初值的敏感，表现出随机性. 解微分方程常用数值方法，记两次采样间的步长记为 h，则方程可由下式逼近：

$$X(t_{n+1}) - X(t_n) = hF(X(t_n)).$$

于是我们又有 $X(t_{n+1}) = f(X(t_n))$，其中 $f = hF + X$. 又回到了离散的动力系统.

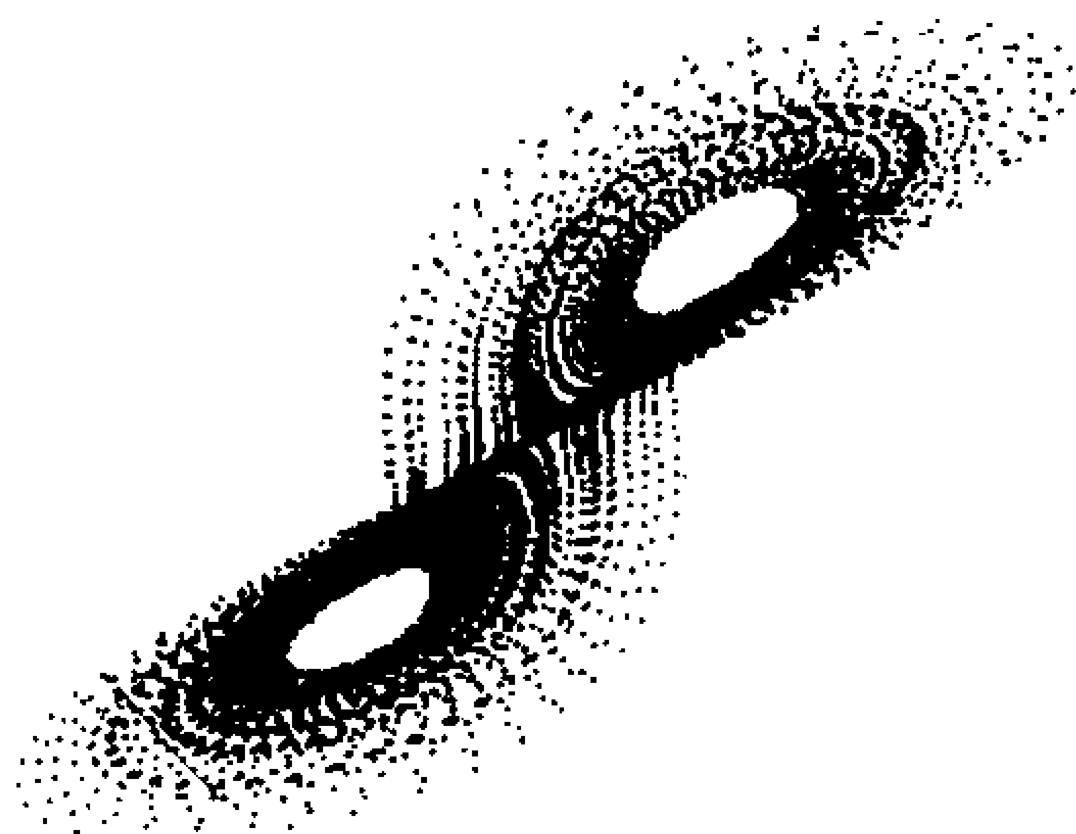
最著名的例子是 Lorenz（1963）所考虑的模拟环球气象的动力系统：

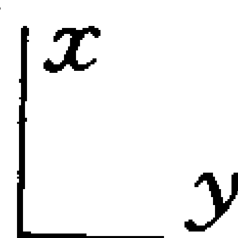
$$\begin{cases} \dfrac{dx}{dt} = \sigma(y - x), \\ \dfrac{dy}{dt} = -xz + rx - y, \\ \dfrac{dz}{dt} = xy - bz. \end{cases}$$

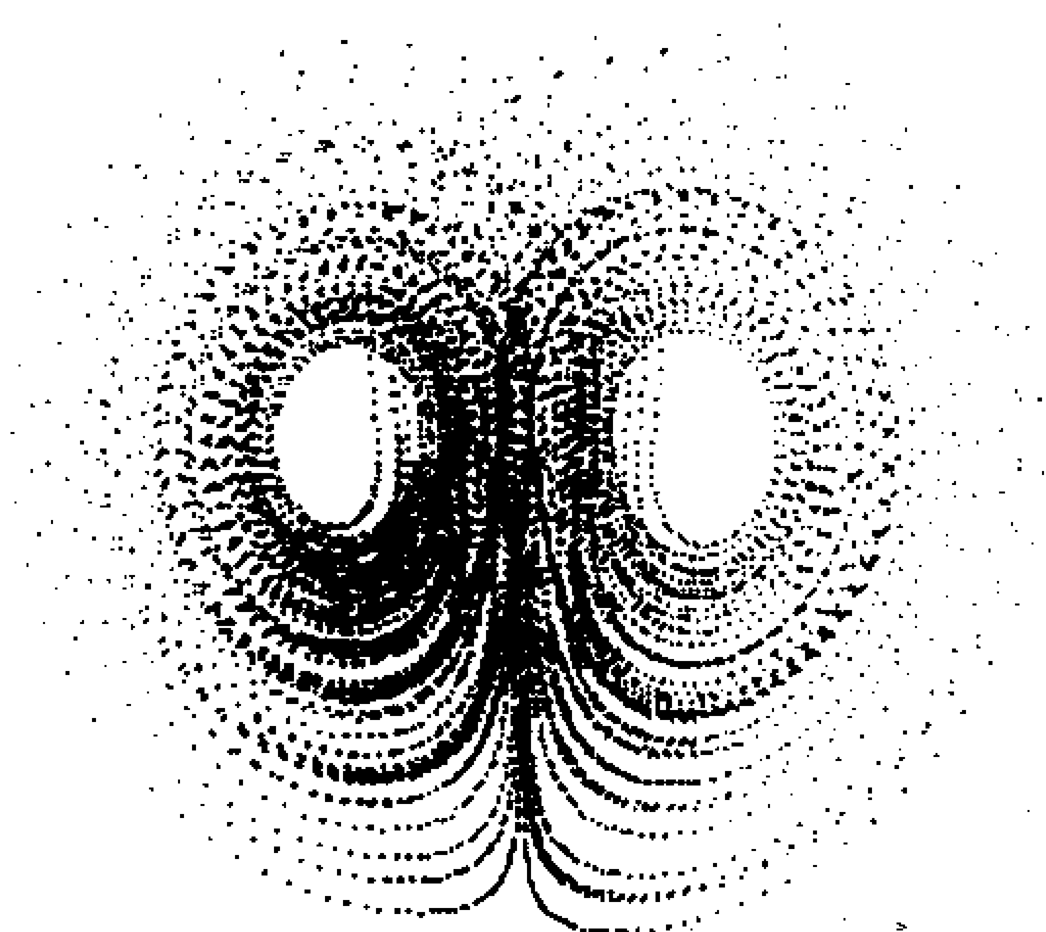
取 $\sigma = 10$，$r = 28$，$b = 8/3$，解轨道表现出对初值的敏感和无法预料的波动. 图 3.6 所呈现的是 Lorenz 吸引子的数值逼近 (Berliner 教授所绘). 多像是一只蝴蝶！方程的真实解为图 3.7 所示. 这是一个稳定的、低维的和非周期的奇异吸引子. 其轨道永不自交. Gleick 说过：混沌的全部丰富内容尽在其中！

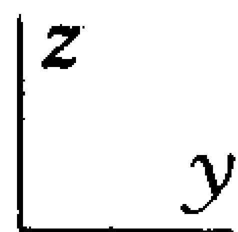
1976 年法国天文学家 Michel Henon 提出一组二维映射：

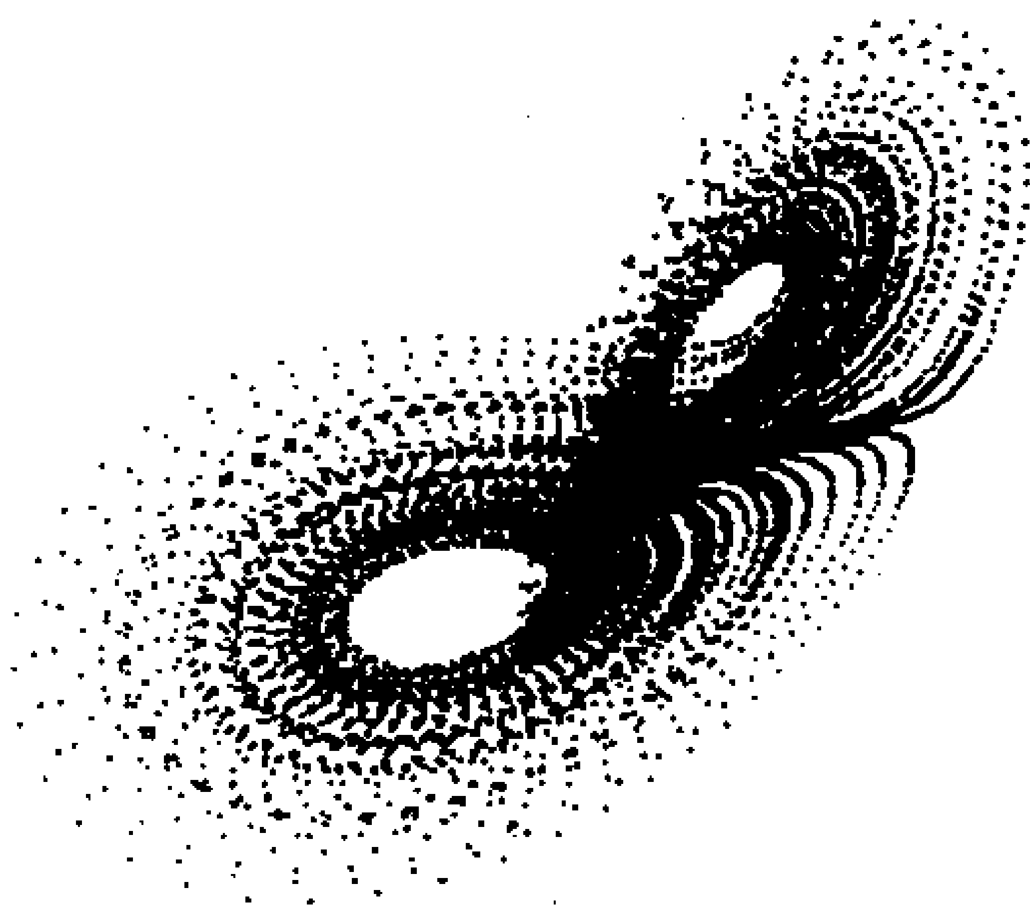
$$H(x,y) = (1 + y - ax^2, bx).$$




x
y




z
y



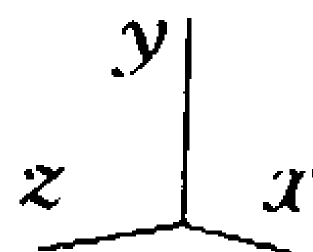

y
z
x



图 3. 6

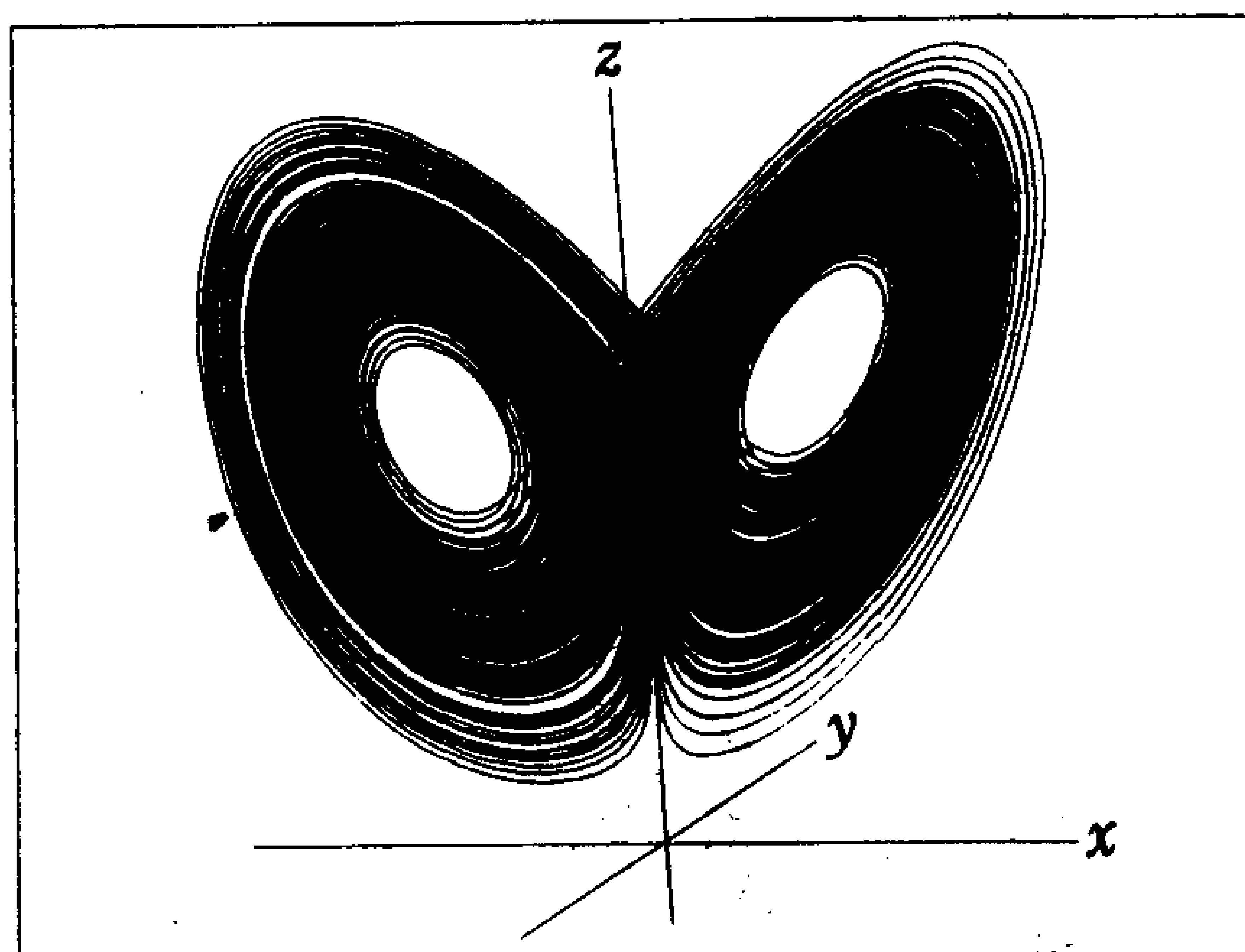


图 3. 7

在迭代过程中 $x_{k+1}=1+y_k-ax_k^2, y_{k+1}=bx_k$. 它由三个变换复合而成：$H(x,y)=H_3\circ H_2\circ H_1\circ(x,y)$，其中 H_1 为在 y 坐标上的非线性弯曲，H_2 为 x 方向的压缩，H_3 为关于对角线的反

射. 取 $a=1.4$，$b=0.3$，随着系统的演化就得到著名的 Henon 吸引子. 所谓吸引子依 Ruelle 的定义是

定义 M 为相空间，f 为一映射，开集 $U\subset M$，若当 t 足够大时对每一个 $V\supset A$ 都有 $f^tU\subset V$，且对所有 t，$f^tA=A$ 时则 A 称为是以 U 为基本邻域的吸引集，开集 $\bigcup\limits_{t>0}(f^t)^{-1}U$ 为 A 的吸附域，小的吸引集又常称作吸引子.

吸引子大体可以分为四类：吸引不动点；吸引周期轨；拟周期吸引子；以及奇异吸引子. 所谓 A 是一个奇异吸引子一般有如下特点：a）对初始值敏感；b）A 是一个吸引子；c）A 有分形几何结构；d）A 不能分成两个不同的吸引子. 对奇异吸引子 A 中的任何点 X 在基本邻域 U 中都存在初始值使其轨道与 X 任意接近. Henon 映射所产生的吸引子如图 3.8 所示正是这样一个奇异吸引子，它也被称为是混沌的. 其分形维数经实验计算为 1.28 左右.

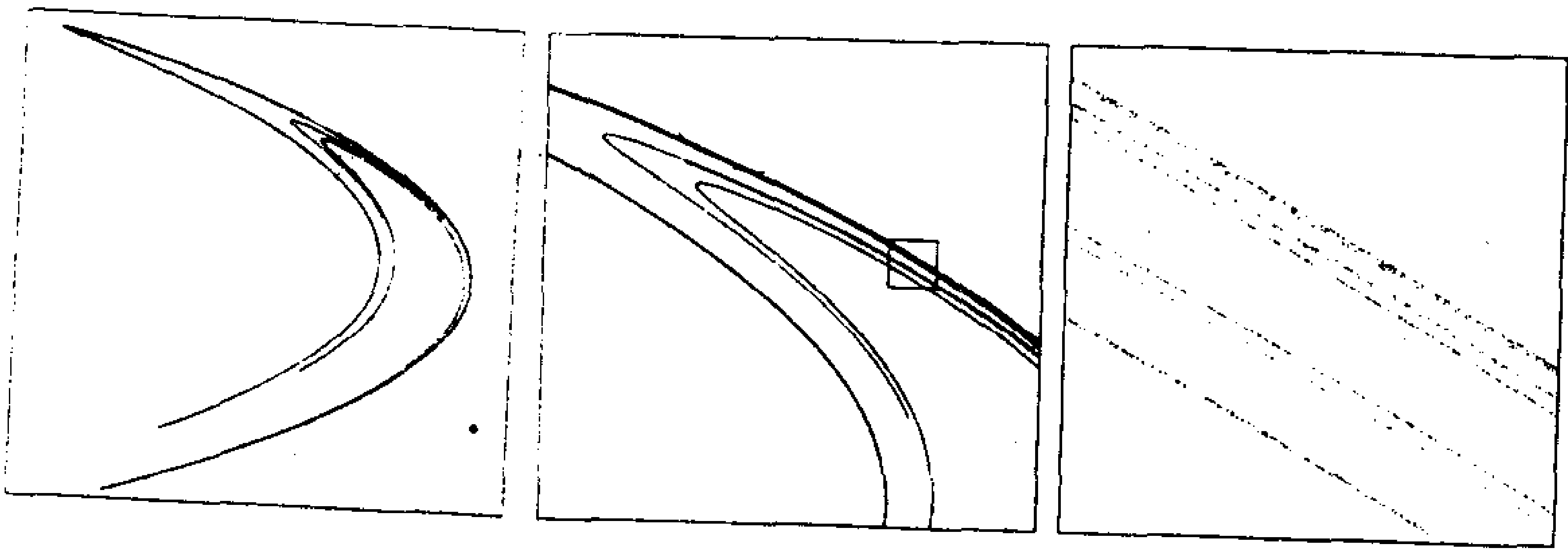


图 3.8

可以形象地讲分形是混沌的签名，不仅奇异吸引子本身呈

分形结构，而吸附域的边界也往往呈现分形结构．为了能数量化地刻画对初始值敏感的奇异吸引子的性状，需要进一步讨论多维 Lyapunov 谱与分形的维数．

粗略而言，在系统随时间演化的过程中初始状态发生小扰动 ΔX_0 时，t 次迭代后系统的变化 ΔX_t 增加或减少的指数速度就是其 Lyapunov 指数．记 $X_t = F(X_{t-1})$ 为 m 维空间中的迭代，F 在 X_t 处的 Jocobian 记为 J_t，$\|\cdot\|$ 为 $m\times m$ 矩阵的模．则当

$$\lim_{t\to\infty}\frac{1}{t}\log\|J_{t-1}\cdot J_{t-2}\cdot\cdots\cdot J_0\Delta X_0\|$$

存在时就是最大的 Lyapunov 指数．以 F 是线性变换为例，若 $X_t = AX_{t-1}$，则 $X_t = A^tX_0$，γ_i 为 A 的特征根（$\det(\lambda I - A) = 0$ 的根）．$\lambda_i = \log|\gamma_i|$，并将 i 调整次序使 $\lambda_1\geqslant\lambda_2>\cdots\geqslant\lambda_m$，则最大的特征根模的对数就是 Lyapunov 指数．当 $\lambda_1>0$ 时，提示了奇异吸引子与混沌．对于非线性迭代，扰动依 Taylor 展开可近似表为 $\Delta X_t\doteq T^t\Delta X_0$，其中 $T^t = J_{t-1}\cdots J_0$．问题是上述极限是否存在？其值将是否依赖于初始值 X_0？好在有多重遍历性定理保证了不依赖于 X_0 的极限存在．（又是遍历性！）

定理 3.3.1（Oseledec，1968）　**多重遍历性定理**　P 为空间 D 上的概率测度，保测变换 f 是遍历的．T：$D\to\{J_{m\times m}$：$m\times m$ 矩阵$\}$ 是可测的（在我们要用的讨论中 T 即切映射，是 f 在 x 处的 Jocobian）且 $E_P[\log^+\|T(X)\|]<\infty$，其中 $\log^+u = 0\vee\log u$．$\|\cdot\|$ 为 m 维空间中的欧氏模．记 $T_x^n = \prod_{k=0}^{n-1}T(f^k(X))$，则关于测度 P 几乎处处有

$$\lim_{n\to\infty}(T_X^{n*}T_X^n)^{\frac{1}{2n}} = \Lambda_X,$$

(其中 T_X^{n*} 为 T_X^n 的伴随矩阵).

Λ_X 的特征根的对数称为 Lyapunov 指数谱，记为 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m$，并由此定义空间

$$E^{(i)} = \{X_0 \in R^m \mid \lambda(X_0) \leqslant \lambda_i\}.$$

则定理 3. 3. 1 的推论是当 $u \in E^{(i)} \backslash E^{(i-1)}$ 时几乎处处有

$$\lim_{t\to\infty} \frac{1}{t} \log \| T^t u \| = \lambda_i.$$

特别当 $u \notin E^{(2)}$，几乎处处成立

$$\lim_{t\to\infty} \frac{1}{t} \log \| T^t u \| = \lambda_1.$$

相对 R^m，$E^{(2)}$是一个低维的子集，因而有了我们所企望的对大部分 x，极限 λ_1 是存在的. λ_i 的几何意义是 $e^{\lambda_1+\cdots+\lambda_k}$ 为 k 维体积元 $\Delta_1 X(t) \wedge \cdots \wedge \Delta_k X(t)$ 的增长比例因子，当 $\Delta X(t)$ 为无穷维的，λ_1就是其指数增长速度. $\sum_{i=1}^{k} \lambda_i < 0$，膨胀变为压缩. $\sum_{i=1}^{k} \lambda_i = 0$ 表示变换保持 k 维体积不变.

例 对 Hénon 映射 $J = \begin{pmatrix} -2a & 1 \\ b & 0 \end{pmatrix}$，$|J| = -b = 0.3$，$\lambda_1 = 0.42$，$\lambda_2 = -1.62$. 奇异性有了量化表示. 对 Lorenz 映射

$$J = \begin{pmatrix} -\sigma & \sigma & 0 \\ r - z & -1 & -x \\ y & x & -b \end{pmatrix}$$

由实验及关于系统轨迹的考量可以得出：$\lambda_1 \doteq 0.9$，$\lambda_2 = 0$，$\lambda_3 \doteq -12.8$. 在某些方向上体积缩小的速度是惊人的：正 Lyapunov 指数反映出奇异性，提示混沌现象；反之，负的 Lyapunov 指数所描绘的是稳定或周期状态.

分形 fractal 一词是分形几何之父 Benoit Mandelbort 创造的，是由拉丁文 fractus（意为破碎、断片）、英文 fraction（分数）启发而得. Mandelbort1924年生于华沙，1936年随家庭迁往巴黎，1958年在美国定居开始其异国科研生涯. 他蜚声全球的两本著作是《构形、机会与维数》，1977及《自然界的分形几何学》，1982. 分形维数是分形数量化的外在表现，分形概念的本质在于其几何特征：这种集合在几乎每一点的领域里点的分布显得疏稠不一、散乱零落；几乎每一点都没有切线；分形集在几何上有某种自相似性. 分形最典型的例子是数学上最有名的 Cantor 集. 它是由 $[0,1]$ 线段不断地依次去掉子区间 $\left(\frac{1}{3},\frac{2}{3}\right)$、$\left(\frac{1}{9},\frac{2}{9}\right)$、$\left(\frac{7}{9},\frac{8}{9}\right)$、…，直至无穷所剩下的集合. 也被叫作 Cantor 尘埃. 它是个闭集但其 Lesbegue 测度为0. 把 Cantor 集一分为二，每一部分又是缩小了 $\frac{1}{3}$ 的完全一样的 Cantor 集！这就是自相似. 在 Mandelbort 最初的定义中分形是其 Hausdorff 维数严格大于拓扑维数的集合（集合的拓扑维数总是整数）. 但这把一些 Hausdorff 维数为整数的明显分形集排除在外. 至今仍没有关于分形的普遍认同的定义，而一般认为称一个集合 F 为分形集应有以下性质：

(i) F 有精细结构，即有任意小比例的细节；

(ii) F 很不规则，其局部与整体不能用传统的几何加以描述；

(iii) F 有自相似的结构；

(iv) F 的分形维数常大于拓扑维数；

(v) F 能以非常简单的方式定义，可能由迭代产生.

最著名的 Mandelbort 集是由复迭代 $z \to z^2+c$ 产生的，此

分形集合如图3.9所示（最后小图为第一图的300 000 000倍）.

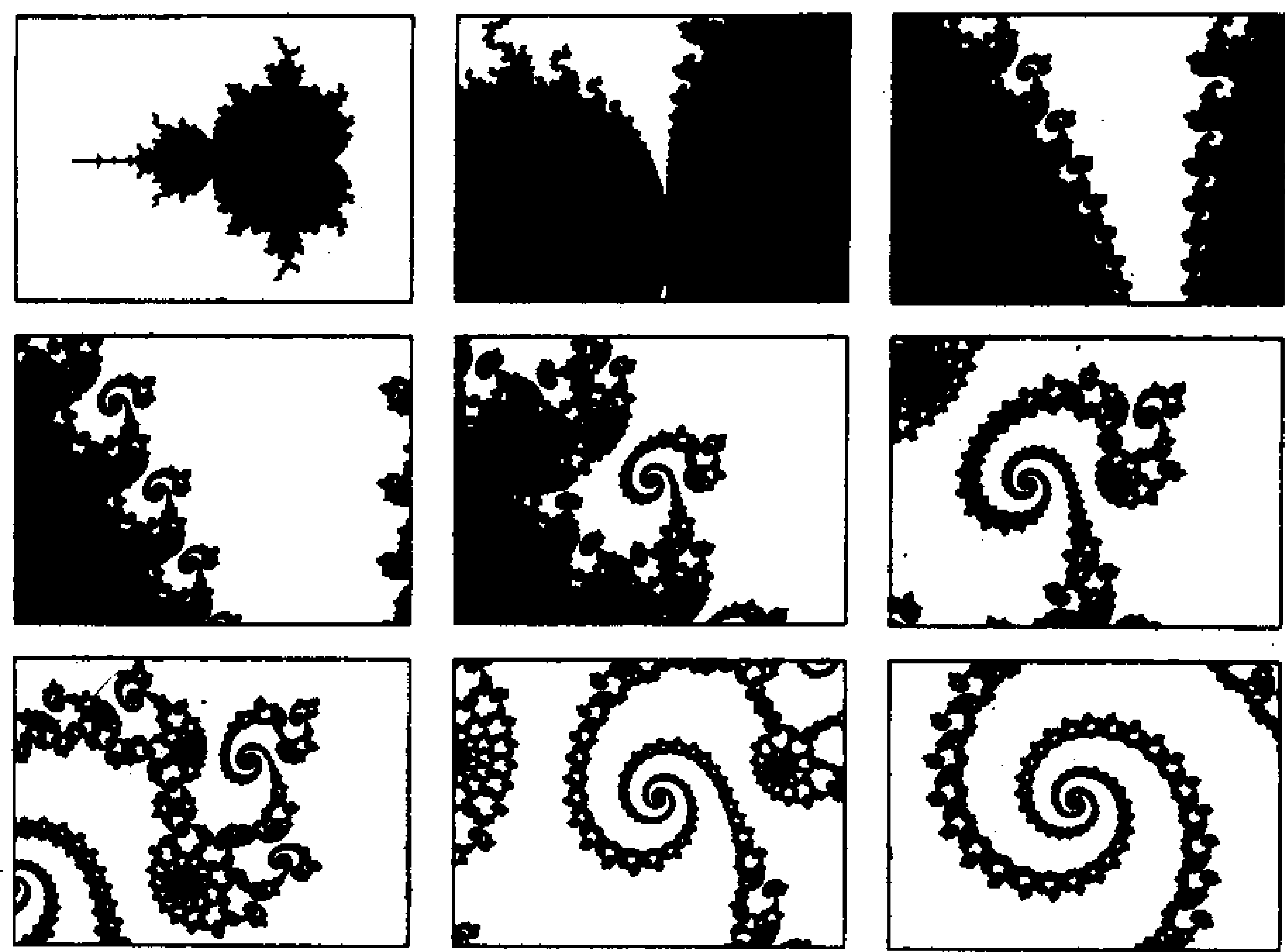


图3.9

由于Cantor集为零测度集，通常欧氏空间的度量方法无法比较分形集的大小. 所以势必要引进分形维数的概念，一个毛线团，远看是一个点算0维，近看是一个3维球，再细看又是由一维的线缠绕而成的有着更精细的结构. 可见讲维数是与尺度有关的，集合的一种不依赖于尺度的内在性质衡量了集合复杂度. 分形集在一定尺度范围内有这种无尺度性，借用尺子丈量长度和面积的想法引入

定义 E 为 $\boldsymbol{R}^n$ 中的非空子集，记 $N_\varepsilon(E)$ 为盖住集合 E 所需的直径为 ε 的闭球的最少个数，则 E 的计盒维数为

$$D_B(E)=\lim_{\varepsilon\to 0}\frac{\log N_\varepsilon(E)}{-\log\varepsilon}.$$

注：上述极限可能不存在，于是可以引入上极限和下极限当两者相等时定义其为 D_B. 容易证明

$$\lim_{\varepsilon\to 0} N_\varepsilon(E)\varepsilon^\alpha = \begin{cases} \infty, & 当\ \alpha < D_B(E), \\ 0, & 当\ \alpha > D_B(E). \end{cases}$$

又若 $E \subseteq \boldsymbol{R}^n$ 有界，$r>2$ 为整数，G_n 为以 r^{-n}为边长的互不相交的一族小立方体，其全体覆盖了 E，则 $N_n(E)$ = 覆盖 E 的 G_n 中小立方体的最少个数仍有

$$D_B(E) = \lim_{n\to\infty} \frac{\log N_n(E)}{\log r^n}.$$

当 $E \subseteq F$ 时，显然有 $D_B(E) \leqslant D_B(F)$，而对两集 E、F，$D_B(E \cup F) = D_B(E) \vee D_B(F)$. 但有限并不能推广到可列个集合的并. 有例子表明 $D_B(\bigcup_k E_k) > \sup_k D_B(E_k)$.

定理3.3.2 E 为 $\boldsymbol{R}^m$ 中的非空有界集合，则当 E 为有限集时，$D_B(E) = 0$，当 E 含有 $\boldsymbol{R}^m$ 中开集时，$D_B(E) = m$；且总有 $D_B(E) \leqslant m$.

证明 记 V 为以 k 为边长的 m 维立方体，$E \subset V$. 取 $r=2$ 则 $N_n(E) \leqslant N_n(V) = k^m 2^{nm}$. 所以 $D_B(E) \leqslant D_B(V) = m$. 而当开集 $B \subseteq E$ 时 B 中含有 m 维立方体，所以 $D_B(E) \geqslant m$. 综合即有 $D_B(E) = m$. 当 E 为有限点集，当 n 充分大 $N_n(E) = k$，故 $D_B(E) = 0$.

例 Cantor 集 C 的计盒维数 $D_B(C)$ 可计算如下. 取 G_n 为形如$\left[\frac{j}{3^n}, \frac{j+1}{3^n}\right]$的3进位网格族，则 $N_n(C) = 2^n$，

$$D_B(C) = \log 2^n / \log 3^n = \log 2 / \log 3 \doteq 0.6309.$$

定义 $S:D\to D$ 称为是 D 上的压缩，是指存在$0<c<1$，对任何 x，$y\in D$ 都有 $\|S(x)-S(y)\|\leqslant c\|x-y\|$ 成立．当等号成立时压缩就称为是相似的，c 称为压缩比．若映射族 S_1，…，S_m：$\boldsymbol{R}^n\to\boldsymbol{R}^n$ 是相似的且各有压缩比 c_i，$i=1$，…，m，则在这映射族下的不变集称为自相似集；满足 $\sum_{i=1}^{m}c_i{}^S=1$ 的 S 称为自相似集 $E=\bigcup_{i=1}^{m}S_i(E)$ 的自相似维数，记为 $D_S(E)$．

例 容易验证对 Cantor 集 C，满足 $\left(\frac{1}{3}\right)^S+\left(\frac{1}{3}\right)^S=1$ 的 S 为 $\frac{\log 2}{\log 3}$．即 $D_S(C)=0.6309$.

定义 令 $E\subseteq R^N$，$\varepsilon>0$，若一组可列个集合 $\{E_k\}$ 满足 $\bigcup_k E_k\supseteq E$，且对任何 k 有 $|E_k|\leqslant\varepsilon<1$，其中 $|\cdot|$ 为集合直径（集合中两点的最大距离），则称 $\{E_k\}$ 为 E 的一个 ε-覆盖.

记，$H^\alpha(E)=\lim\limits_{\varepsilon\to 0}\left\{\sum\limits_k|E_k|^\alpha\middle|\{E_k\}\text{ 为 }E\text{ 的 }\varepsilon\text{- 覆盖}\right\}$.

当 $|E_k|\leqslant\varepsilon<1$，$|E_k|^\alpha$ 是 α 的增函数，于是 $H^\alpha(E)$ 为 α 的减函数，而且一旦有 α 使 $H^\alpha(E)<\infty$时，则对所有 $\beta>\alpha$ 有 $H^\beta(E)=0$（参看计盒维数定义的注）．于是又可以

定义 对 $E\subseteq\boldsymbol{R}^n$，满足当 $\alpha<\alpha(E)$ 时 $H^\alpha(E)=\infty$，$\alpha>\alpha(E)$ 时 $H^\alpha(E)=0$的唯一实数称为 E 的 Hausdorff 维数，记为 $D_H(E)$，也即

$$D_H(E)=\inf\{\alpha>0\,|\,H^\alpha(E)=0\}.$$

注 $D_H\left(\bigcup_k E_k\right)=\sup_k D_H(E_k)$；对 R^n 之 E 恒有 $D_H(E)\leqslant D_B(E)\leqslant n$.

Hausdorff 维数与计盒维数都涉及覆盖，但前者可用大小不同的集合覆盖，而后者用以覆盖的集合直径一样，所以 D_H 似乎更准确一些. 也有文献研究从内部填充的填充维数. 从注可知 D_H 在满足可列并的等式方面是对 D_B 的改进. D_B 为 D_H 的上界. Hausdorff 维数的计算也并不容易，但有下面有用的定理

定理3.3.3 $S_1,\cdots,S_m$ 为相似映射，若存在非空的有界开集 V 使 $\bigcup_{i=1}^{m}S_i(V)\subset V$ 及 $E=\bigcup_{i=1}^{m}S_i(E)$，则

$$D_H(E)=D_B(E)=D_S(E).$$

证明可参看 Kenneth Falconer 的《分形几何》一书.

例 Koch 曲线是自相似分形的典型例子，它是由极简单的迭代机制产生的；以一段直线为初始值，将它缩短$\frac{1}{3}$后用四段拼接起来，每次迭代为

$$S(A)=S_1(A)\cup S_2(A)\cup S_3(A)\cup S_4(A).$$

其中 $S_1:(x,y)\to\left(\frac{1}{3}x,\frac{1}{3}y\right)$,

$$S_2:(x,y)\to\left(\frac{1}{6}x-\frac{\sqrt{3}}{6}y+\frac{1}{3},\frac{\sqrt{3}}{6}y+\frac{1}{6}y\right),$$

$$S_3:(x,y)\to\left(\frac{x}{6}+\frac{\sqrt{3}}{6}y+\frac{1}{2},\right.$$

$$\left.-\frac{\sqrt{3}}{6}x+\frac{1}{6}y+\frac{\sqrt{3}}{6}\right),$$

$$S_4:(x,y)\to\left(\frac{1}{3}x+\frac{2}{3},\frac{1}{3}y\right).$$

这是瑞典数学家 Helge von Koch1904年引入的，也称作雪花曲线．形如图3.10.

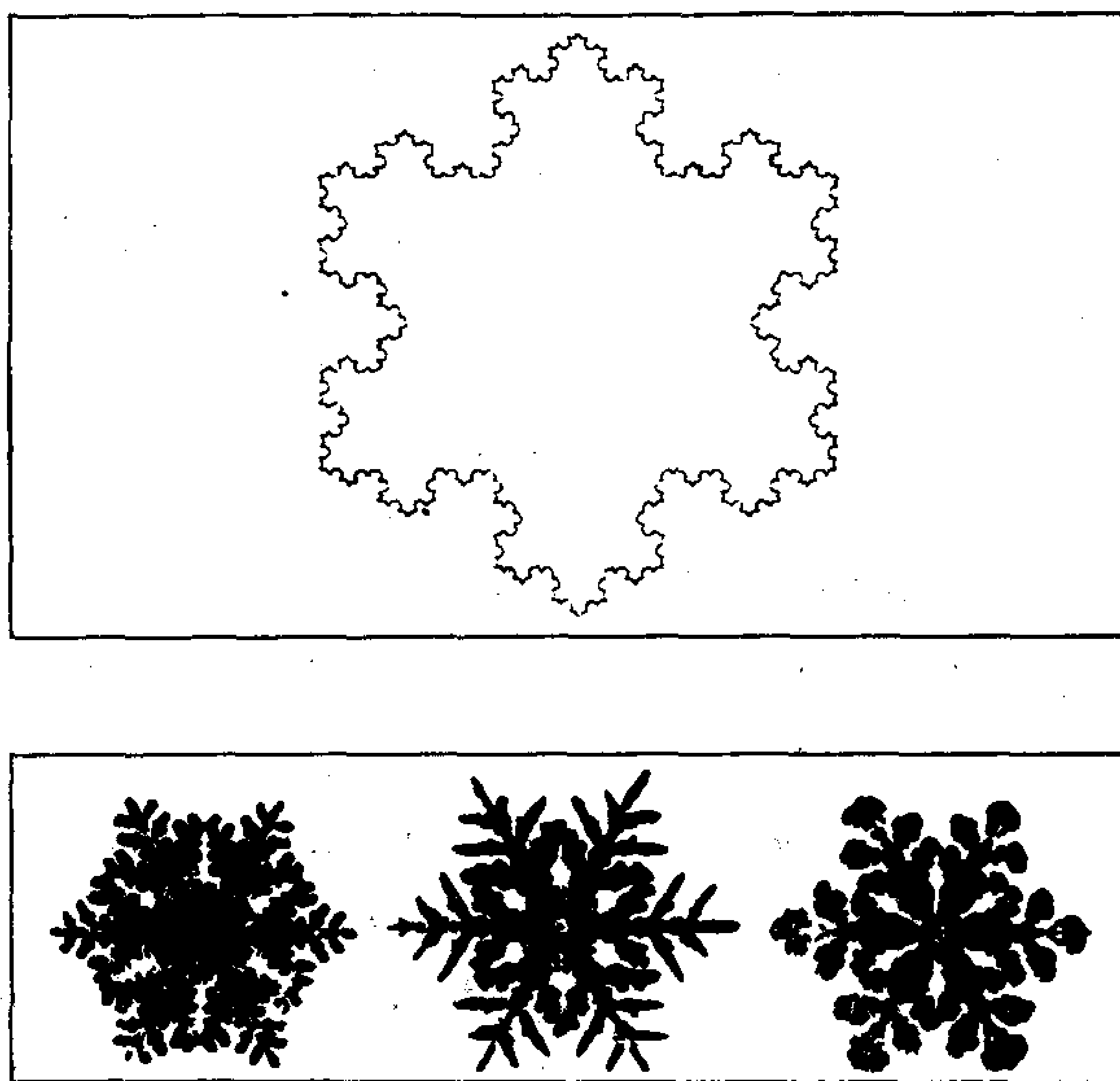


图 3.10

读者可以将它与现实生活中的雪花作比较．它的相似维数为 $\log4/\log3\doteq1.2619$.

例 除确定性集合之外，由随机过程 $X(t)$ 的实现所产生的值域集合 $\boldsymbol{R}=\{X(t,\omega)\,|\,t\in T\}$，水平集合 $L=\{t\in T\,|\,X(t,\omega)=l\}$ 及图集 $G=\{(t,X(t,\omega))\,|\,t\in T\}$ 的分形性质是统计学者所关心的．对 Brown 运动 $B(t)$ 有 $D_H(R)=2, D_H(L)=\frac{1}{2}$, $D_H(G)=1.5$，从而表现出 Brown 运动轨道的不规则特性.

从统计学的眼光看，在混沌所产生的奇异吸引子上演化过程访问吸引子上各点的次数是不一样多的．于是有必要考虑其上的概率分布．为使讨论简化，仅考虑 $\boldsymbol{R}^n$ 上定义的概率测度 μ，$\mu(E), E \in R^n$ 代表集合 E 上的概率．使 $\mu(\cdot) > 0$ 的集合称为 μ 的支撑．比如一维正态分布的支撑就是 $\boldsymbol{R}$，而 $U[0,1]$ 的支撑为 $[0, 1]$ 区间等等．若 μ 是 $\boldsymbol{R}^n$ 上的绝对连续分布，则 $\mu(E) > 0$ 意味着 E 的 Lebesgue 测度为正，所以由定理3.3.2知 $D_H(E) = n$．此时称 μ 为 n 维的．但在一般复杂的情形，μ 将在 $[0, n]$ 的 Borel 子集上有一个维数分布．

定义 记 $\mathscr{B}(R^n)$ 为 R^n 中的 Borel 子集全体，μ 在区间 $[0, \alpha]$，$\alpha \leqslant n$ 上的 Hausdorff 维数分布为

$$\mu_H([0,\alpha]) = \sup\{\mu(B) \mid D_H(B) \leqslant \alpha, B \in \mathscr{B}(R^n)\}.$$

有了此维数分布就可以把 D_H 看成服从 μ_H 分布的随机变量，于是可以定义平均 Hausdorff 维数为 $E_{\mu_H}(D_H)$．当 μ_H 为单点分布时，表明存在 Borel 集 B，其 $D_H(B) = \alpha$，使 $\mu(B) = 1$，而对任何 E 当 $D_H(E) < \alpha$ 时 $\mu(E) = 0$. 这时 $D_H(B) = \alpha$ 才是精确的．下面定义逐点维数，记 $B(x,\varepsilon)$ 为以 x 为中心，以 ε 为半径的小球，则有

定义 测度 μ 在点 x 的逐点维数 $\alpha_\mu(x)$ 为

$$\alpha_\mu(x) = \lim_{\varepsilon \to 0} \frac{\log\mu(B(x,\varepsilon))}{\log\varepsilon}.$$

$\alpha_\mu(x)$ 在文献中也被称为是局部 Hölder 指数.

$\alpha_\mu(x)$ 与 $\mu_H([0,\alpha])$ 的相互关系为

定理3.3.4 沿用上述记号，对每一个 $\alpha \geqslant 0$ 记 $K_\alpha = \{x \mid \alpha_\mu(x) \leqslant \alpha\}$，则

(i) $D_H(K_\alpha) \leqslant \alpha$;

(ii) $\mu(K_\alpha) = \mu_H([0,\alpha])$.

证明 设有一串不交的闭球 $\{B_k\}$ 覆盖 K_α 且满足 $\varepsilon > |B_k| = 2\mu(B_k)^{\frac{1}{\alpha_k}}$. 其中 $\alpha_k \leqslant \alpha$. 所以，令 β 为任取的 $>\alpha$ 的数，则

$$\sum_k |B_k|^\beta = 2^\beta \sum_k \mu(B_k)^{\frac{\beta}{\alpha_k}} \leqslant 2^\beta \sum_k \mu(B_k)^{\frac{\beta}{\alpha}}$$

$$\leqslant 2^\beta \sum_k \mu(B_k) \leqslant 2^\beta.$$

因此，$H^\beta(K_\alpha) \leqslant 2^\beta < \infty$，故有 $D_H(K_\alpha) \leqslant \alpha$.

(i) 得证.

由定义 $\mu(K_\alpha) \leqslant \mu_H([0.\alpha])$，再证另一方向即得所欲证.

定理的一个重要推论是当 μ 有精确的 Hausdorff 维数 α 时，关于测度 μ 几乎处处有 $\alpha_\mu(x) = \alpha$. 这使我们在 μ 未知时基于数据来估计 D_H 成为可能. 如果说逐点维数仅着眼于局部情形，那么信息维数和相关维数则基于对疏密不一的集合的整体考量. 有了概率测度 μ，记 $B(x,\varepsilon)$ 的测度为 $\mu_\varepsilon(x)$，则可有

定义 μ_ε 的 L^q 模定义为

$$\|\mu_\varepsilon\|_q = (E(\mu_\varepsilon(x))^q)^{\frac{1}{q}}.$$

这是距离与变差概念的推广与抽象. 基于此又有：

定义 μ 的 q 阶 Renyi 信息维数为

$$D_{R,q}=\lim_{\varepsilon\to 0}\frac{\log\|\mu_\varepsilon\|_q}{\log\varepsilon}.$$

定义 记 $A_\varepsilon=\{(x,y)\mid\|x-y\|\leqslant\varepsilon\}$，关于测度 μ，记 $C_\mu(\varepsilon)=\mu\times\mu(A_\varepsilon)$.（文献中称为相关积分，$\mu\times\mu$ 为乘积测度，即重积分 $\iint 1_{A_\varepsilon}\mu(dx)\mu(dy)$.）

$$D_C=\lim_{\varepsilon\to 0}\frac{\log C_\mu(\varepsilon)}{\log\varepsilon},$$

称为是 μ 的相关维数.

定义 μ 的信息维数定义为

$$D_I=\lim_{\varepsilon\to 0}\frac{E(\log\mu(B(x,\varepsilon)))}{\log\varepsilon}.$$

易见 $D_{R,1}=D_C$，这是因为

$$C_\mu(\varepsilon)=\iint 1_{A_\varepsilon}\mu(dx)\mu(dy)=E\mu(B(x,\varepsilon))=\|\mu_\varepsilon\|_1.$$

而 $\lim\limits_{q\to 0}D_{R,q}=D_I$，这可由

$$\lim_{q\to 0}(E(\mu(B(x,\varepsilon))^q))^{\frac{1}{q}}=E[\log\mu(B(x,\varepsilon))]$$

得知. 在实际中 $C_\mu(\varepsilon)$ 可用基于观察值的 $C_N(\varepsilon)$ 来逼近，其中

$$C_N(\varepsilon)=\frac{1}{N^2}\#\{(x,y)\mid\|x-y\|\leqslant\varepsilon\}.$$

N 代表吸引子上全部观察数，分子为距离小于 ε 的观察点对数. 若吸引集可由 $N(\varepsilon)$ 个小球所覆盖，记 π_i 为 μ 在第 i 个小体积元上的概率，并记 $H(\varepsilon)=-\sum\limits_{i=1}^{N(\varepsilon)}\pi_i\log\pi_i$，则 $H(\varepsilon)$ 是 $E[\log\mu(B(x,\varepsilon))]$ 的近似. 于是 $D_I\doteq\lim\limits_{\varepsilon\to 0}\dfrac{\log H(\varepsilon)}{\log\varepsilon}$. 信息维数

也因 $H(\cdot)$ 与熵的联系而得名. 人们自然也对不同维数定义之间的关系感兴趣，由凸函数的性质及 Jensen 不等式可以证明：

$$D_B \geqslant D_H \geqslant D_I \geqslant D_C.$$

“分形的相关维数”是物理文献中出现的说法，有时会误导读者. 这一说法没能区分分形与分形集中的概率分布的不同. 两者的研究都能提供有关动力系统演化中结构复杂性、不均匀性方面的信息. 混沌动力系统的遍历测度维数的正则性使逐点维数关于 μ 几乎处处为常数，且等于支撑 μ 所需最小集合的 Hausdorff 维数. 正是这一点使人们有可能基于物理实验观察数据来估计维数. 在 $\alpha_\mu(x)\equiv\alpha$ 的情形称吸引子为齐次分形，而当 $\alpha_\mu(x)$ 依赖 x，μ 有一个分布时称吸引子为多重分形，其研究日趋活跃. (英文为 Multifractal.)

从本节引入的 Lyapunov 指数 $\lambda_1>\lambda_2>\cdots$ 出发还可引入

定义 记 $j=\max\{k\,|\,\lambda_1+\cdots+\lambda_k\geqslant 0\}$，
则

$$D_L = j+\frac{\lambda_1+\cdots+\lambda_j}{|\lambda_{j+1}|}$$

称为是映射 f 的 Lyapunov 维数.

D_L 与 f 所产生的吸引子的维数之间也有本质的关系. 1981 年 F. Ledrappier 证明了 $D_L\geqslant D_H$，1979年 Kaplan 和 York 曾猜想 $D_L=D_H$. 1982年 Young 对二维情形证明了若 μ 为映射 f 的不变测度，且 $\lambda_1>\lambda_2$ 则 $D_H=h_\mu(f)\left(\frac{1}{\lambda_1}-\frac{1}{\lambda_2}\right)$，其中 $h_\mu(f)$ 为 f 关于 μ 的 Kolmogrov 熵. 当 $\lambda_1>0>\lambda_2$时 $h_\mu(f)=\lambda_1$，于是 $D_H=\lambda_1\left(\frac{1}{\lambda_1}-\frac{1}{\lambda_2}\right)=1+\frac{\lambda_1}{|\lambda_2|}=D_L$. 但1983年，Grassbeger 和

定义 函数 $H(P_1,\cdots,P_n)$ 称为是 $(\Omega, \mathscr{F}, P)$（其中 $P=\{P_i, i=1, \cdots, n\}$）的 Shannon 熵.

定理3.4.1 对确定的 n，等概率分布比不等概率分布有更大的熵，即

$$H\left(\frac{1}{n},\frac{1}{n},\cdots,\frac{1}{n}\right)$$
$$=\max\left\{H(P_1,\cdots,P_n),\sum P_i=1,P_i\geqslant 0\right\}.$$

证明 由 $\log u$ 是严格凹函数，先设 $P_i>0$，故 $-\sum_{i=1}^{n}P_i\log P_i$

$$=\sum_{i=1}^{n}P_i\log\frac{1}{P_i}\leqslant\log\left(\sum_{i=1}^{n}\frac{P_i}{P_i}\right)=\log n=H\left(\frac{1}{n},\cdots,\frac{1}{n}\right);$$

若有 $P_1=0$，则 $H(0,P_2,\cdots,P_n)\leqslant H\left(\frac{1}{n-1},\cdots,\frac{1}{n-1}\right)<H\left(\frac{1}{n},\cdots,\frac{1}{n}\right)$. □

2. Kolmogrov 熵与拓扑熵

对§3.1中的锯齿变换 $S(x)=2x(\bmod 1)$，$\Omega=[0,1]$，S：$\Omega\to\Omega$ 的不变测度为 $U[0,1]$ 也即 Lebesgue 测度 m. 令 $\xi=\left\{\left[0,\frac{1}{2}\right],\left[\frac{1}{2},1\right]\right\}$，则 $S^{-1}(\xi)=\left\{\left[\frac{i-1}{4},\frac{i}{4}\right],i=1,\cdots 4\right\}$（单点测度为0不影响下面讨论），且有

$$\bigvee_{i=0}^{n-1}S^{-i}(\xi)=\left\{\left[\frac{i-1}{2^n},\frac{i}{2^n}\right],i=1,\cdots,2^n.\right\}.$$

若已知 $x,S(x),S^2(x),\cdots,S^{n-1}(x)$ 的信息定出 $S^n(x)$ 的条件不确定度，并考虑其当 $n\to\infty$ 时的极限就是映射 S 关于 ξ 的 Kolmogrov 熵的大致含义. 确而言之，记 $\xi=(A_1,\cdots,A_k)$ 为 Ω 之

一有限分划. $f^{-1}(\xi)=(f^{-1}(A_1),\cdots,f^{-1}(A_k))$ 也是 Ω 的一有限分划. 记

$$\xi^{(n)}=\xi\vee f^{-1}(\xi)\vee\cdots\vee f^{-n+1}(\xi)=\bigvee_{i=0}^{n-1}f^{-i}(\xi),$$

(因为 f 有可能是多对一的映射，所以取 f^{-1})，$\xi^{(n)}$ 为映射在长为 n 的时间段内生成的分划. 已知 ξ, f^{-1} (ξ) 的条件不确定度为 $H(f^{-1}(\xi)|\xi)$. 依 Shannon 熵（记 μ 为 (Ω, $\mathscr{F}$) 上的测度）

$$\begin{aligned}&H(f^{-1}(\xi)|\xi)\\&=\sum_{i=1}^{k}\mu(A_i)\Big(-\sum_{j=1}^{k}\mu(f^{-1}(A_j)|A_i)\log\mu(f^{-1}(A_j)|A_i)\Big)\end{aligned}$$

由熵的性质(3)，

$$H(f^{-1}(\xi)|\xi)=H(\xi\vee f^{-1}(\xi))-H(\xi).$$

类似可推出

$$H(f^{-n}(\xi)|\xi^{(n)})=H(\xi^{(n+1)})-H(\xi^{(n)}).$$

定义 ξ 为 Ω 的有限分划；f: $\Omega\to\Omega$ 为可测变换，f 关于 ξ 的熵定义为

$$h_\mu(f,\xi)=\lim_{n\to\infty}H(f^{-n}(\xi)|\xi^{(n)}),$$

而 $$h_\mu(f)=\sup_{\xi}\{h_\mu(f,\xi):\xi\text{ 为 }\Omega\text{ 之有限分划}\},$$

即为 f 之 Kolmogrov 熵.

式中的上确界是对所有有限分划取的.

定理3.4.2 若 f 为 (Ω, $\mathscr{F}$, μ) 上的保测变换，则对 Ω 的任一有限分划 ξ 有

$$\lim_{n\to\infty}H(f^{-n}(\xi)|\xi^{(n)})=\lim_{n\to\infty}\frac{1}{n}H(\xi^{(n)}).$$

证明 由 f 的保测性知 $\mu(f^{-1}(A))=\mu(A)$，于是亦有 $H(f^{-1}(\xi))=H(\xi)$. 仍由 (3)

$$H(f^{-1}(\xi)|\xi)=H(f^{-1}(\xi)\vee\xi)-H(\xi)$$
$$=H(f^{-1}(\xi)\vee\xi)-H(f^{-1}(\xi))$$
$$=H(\xi|f^{-1}(\xi)),$$

于是由归纳法易证

$$H(f^{-n}(\xi)|\xi^{(n)})=H(\xi|\xi^{(n+1)}).$$

由 $\xi^{(n)}\geqslant\xi^{(n+1)}$ 知 $H(f^{-n}(\xi)|\xi^{(n)})$ 关于 n 单调，故极限存在，反复利用(3) 知

$$H(\xi^{(n)})=H(\xi)+\sum_{i=1}^{n-1}H(f^{-1}(\xi)|\xi^{(i)})$$
$$=\sum_{i=0}^{n-1}H(f^{-1}(\xi)|\xi^{(i)})$$
$$=\sum_{i=0}^{n-1}H(\xi|\xi^{(i)}),$$

于是

$$\lim_{n\to\infty}\frac{1}{n}H(\xi^{(n)})=\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}H(\xi|\xi^{(i)})$$
$$=\lim_{n\to\infty}H(\xi|\xi^{(n)})=\lim_{n\to\infty}H(f^{-n}(\xi)|\xi^{(n)}).\qquad\square$$

定理给出了与定义等价的一个条件，它的直观含义是在映射不断迭代的过程中的平均信息再创造 $\frac{1}{n}H(\xi^{(n)})$. 另一个重要事实是 Kolmogrov 熵是共轭保测变换的不变量.

在一般的拓扑空间中并没有测度的概念. 但当 X 为一个紧的 Hansdorff 空间时，X 有有限的开覆盖 $\mathscr{A}=\{A_1,\cdots,A_k\}$ (想象一下闭区间可以用有限个开集覆盖). 若 $\mathscr{A}$ 中的子集合族也能覆盖 X 则称为是其子覆盖，$\mathscr{A}$ 中开集数最少的子覆盖

称为是极小子覆盖. 极小子覆盖中元素的个数记为 $N(\mathscr{A})$，并赋其中每一开集以相同的概率 $\frac{1}{N(\mathscr{A})}$. 则可知其 Shannon 熵为

$$-\sum \frac{1}{N(\mathscr{A})}\log \frac{1}{N(\mathscr{A})} = \log N(\mathscr{A}),$$

记为 $H(\mathscr{A})$. 与 $\xi \vee \eta$ 类似定义 $\mathscr{A} \vee \mathscr{B}$（$\mathscr{B}$ 为另一开覆盖），则有

$$N(\mathscr{A} \vee \mathscr{B}) \leqslant N(\mathscr{A})N(\mathscr{B}).$$

因而 $H(\mathscr{A} \vee \mathscr{B}) \leqslant H(\mathscr{A}) + H(\mathscr{B})$. 记 $N(\mathscr{A}, f, n)$ 为覆盖 $\bigvee_{i=0}^{n-1} f^{-i}(\mathscr{A})$ 中极小子覆盖的元素个数，则

定义 连续映射 f 关于有限子覆盖 $\mathscr{A}$ 的拓扑熵为 $h_t(f,\mathscr{A}) = \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} f^{-1}(\mathscr{A}))$，据此又定义紧 Hausdorff 拓扑空间 X 上的连续映射 f 的拓扑熵为

$$h_t(f) = \sup\{h_t(f,\mathscr{A}) : \mathscr{A} \text{ 为 } X \text{ 之有限开覆盖}\}.$$

拓扑熵与 Kolmogrov 熵的关系有如下：

定理3.4.3 f 为紧拓扑空间 X 上的连续映射，则

$$h_t(f) = \sup\{h_\mu(f) : \text{测度 } \mu \text{ 关于 } f \text{ 遍历}\}.$$

Ruelle 在1978年还得到一个建立 Kolmogrov 熵与 Lyapunov 指数之间关系的重要结果：若 f 为有限维流型上的可微映射，μ 为遍历测度，则

$$h_\mu(f) \leqslant \sum_{i:\lambda_i>0} \lambda_i$$

这期间 Pesin (1977)、F. Ledrappier (1985) 等人在一定条件下建立了

$$h_\mu(f) = \int \sum_{i:\lambda_i(x)>0} \lambda_i(x) m_i(x) d\mu(x)$$

的充分必要条件. 称为 Pesin 公式. 其中 $\lambda_i(x)$ 为在点 x 处的 Lyapunov 指数, $m_i(x)$ 为特征根 $\lambda_i(x)$ 的重数. $\mu(\cdot)$ 为不变测度. 这些结果表明正 Lyapunov 指数之和在某种意义下与关于相应测度 μ 的 Kolmogrov 熵是相等的, 而定理3.4.3实际上也告诉我们 $h_t(f) \leqslant \sum_{i:\lambda_i>0} \lambda_i$, 因此拓扑熵大于0意味着 f 有正 Lyapunov 指数提示了混沌!

在观察系统演化的进程中, 人们将获得越来越多的关于初始值何在的信息. 因此混沌的道路也是"创造信息"的过程. 由于 n 次迭代后偏差 $\Delta X_n \doteq e^{\lambda n} \Delta X_0$, 假定把 $f^n(X_0)$ 将要访问的区域分成 $[e^{\lambda n}]$ 个小单元, $f^n(X_0)$ 落入每个单元的概率将大约等于 $\frac{1}{[e^{\lambda n}]}$, 其 Shannòn 熵为 $-\sum \frac{1}{[e^{\lambda n}]} \log \frac{1}{[e^{\lambda n}]} = \lambda n$, 当 $\lambda > 0$ 时, 不确定性或信息量将迅速增长. 从这个意义上讲混沌是信息的源泉, 走向混沌是信息的极度膨胀.

最后, 顺便提一下 Boltzman 熵的定义: 对于密度函数 $p(x)$, $H(p) = -\int p(x) \log p(x) dx$. 可以证明在随机变量均值为0, 方差为1的条件下, 标准正态分布有最大的 Boltzman 熵. $H(\phi) = \log \sqrt{2\pi} + \frac{1}{2}$. 这是因为正态分布表现出由大量随机因素造成的最大不确定性.

第四章　走出混沌

科学不只是为了解释现象，更不只是为了说明一些事情，科学的主要任务是建立数学模型.

—Von Neumann

混沌现象是系统的复杂无序和行为的不可预料的同义语，初始值的微小误差可以把事情弄得一塌糊涂. 在一个完全确定的简单系统中可以表现出宛如随机的现象. 而我们通常的以线性模型和 Gauss 分布为主要工具的统计方法竟无法检测出二者的区别，这不能不使传统的统计学面临挑战. 挑战与机遇并存. 确定性系统的这种与随机性的联系使统计理论与方法在混沌现象的研究和混沌实验数据的分析中作用巨大. 面对挑战统计学首先要回答的问题是什么是随机性?如何将混沌与随机两者加以区别?统计学主要是通过建立恰当的模型来分析带误差的实测数据，寻找出隐含的演化规律，特别是简单、低维的结构以寻求自然现象的各种奥秘. 进而，对现象作出解释，有可能的情况下要进行预测. 对此，非线性时间序列模型将成为一

种有力工具. 统计学还能对混沌的重要数字特征: Lyapunov 指数, 吸引子的分形维数进行统计估计. 用以作为其他学科研究混沌的佐证. 探求非线性迭代产生的动力系统的遍历性条件以及寻求映射 f 的不变测度也都属于概率论研究的对象. 近年来活跃的多重分形, 随机瀑布, 随机分形更都离不开深刻的概率统计背景. 混沌科学所提出的问题够统计学家们忙几十年的. 统计学也将为揭开混沌科学的奥秘作出贡献, 提供一条走出混沌之路.

§4.1 统计与混沌

1. 随机性及其检验

Spinoza 曾说过: 自然界中没有什么事物是随机的, 仅仅因为人们对它的了解不完全它才表现出随机性. 扔钱币要算是最典型的随机试验, 引入概率的概念几乎离不开以扔钱币作例子. Stanford 大学的统计教授 Persi Diaconis 在一次访谈中向已故的统计学家 Degroot 展示了图4.1. 图中 ω 为每秒钟钱币的翻转次数, g 为重力常数, u 为扔币的初速度, 密密麻麻的双曲线代表了正面与反面的边界. 按 Diaconis的观点扔币本没有任何随机性, 知道了向上的初速度及转动的速率, 依牛顿定律, 钱币落地时是正还是反是完全确定的物理现象, 何"随机"之有?! Keller 帮助他作了物理实验, 并将初始面与终止面相同的实验结果画在同一个区域内, 相邻的区域的扔币结果则是正、反相导的. 从图上可以看到随着初速度和旋转速率的加快两相邻区域越来越靠近. 于是初始条件的微小变化就能改变结果, 使预测非常之困难. 这就揭示了扔钱币呈现随机性的原 由. 同样, 确

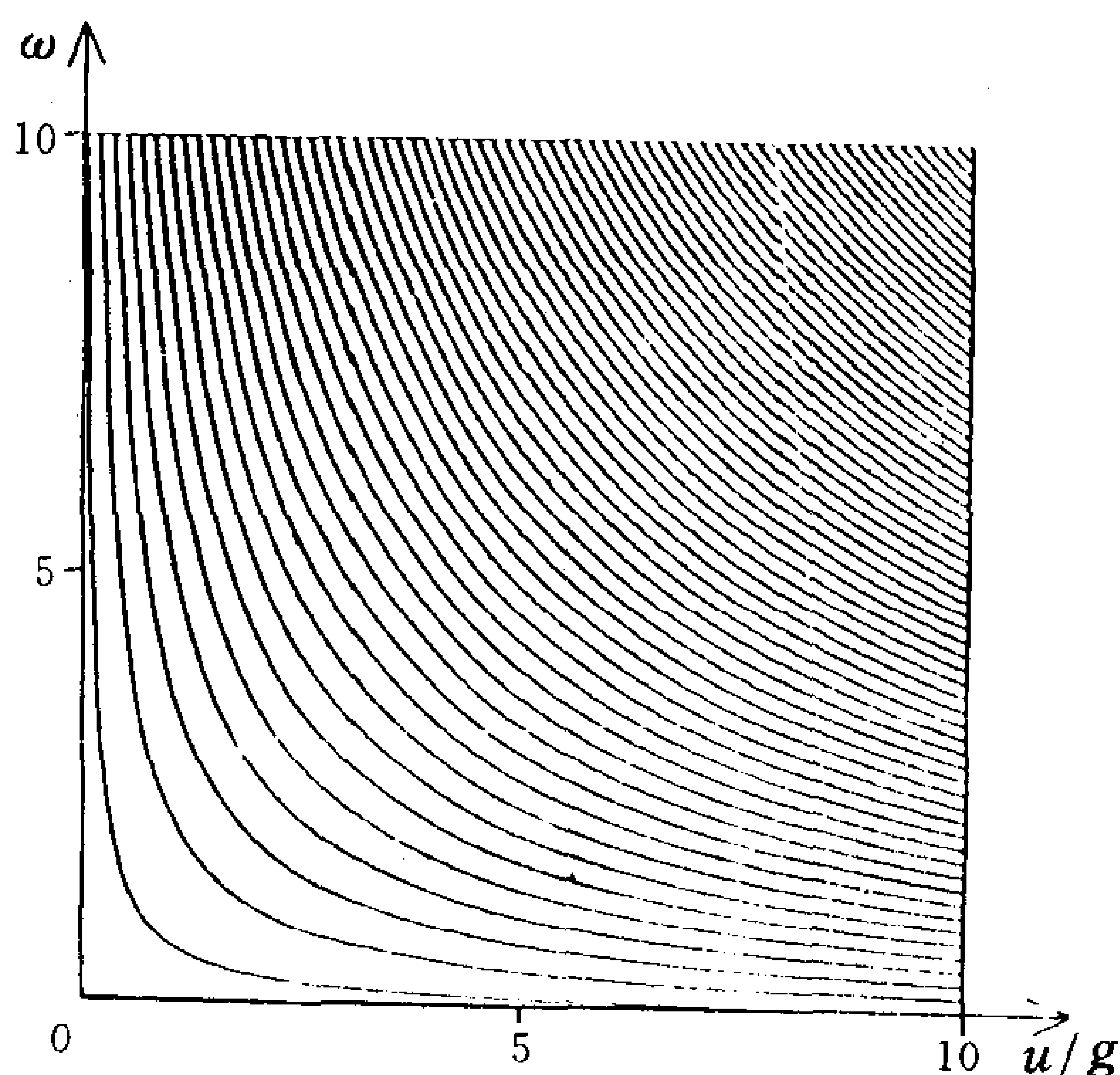


图4.1

定性的 Logistic 迭代也对初值敏感并呈现随机行为.

在混沌的研究中人们大量地使用计算机实验. 无论你所使用的计算机的精度有多高却只能运算有限位小数. 它仅仅是实际数据的一个近似. 计算机所产生的随机数只是伪随机数. 设计好的算法来寻求随机性更好的随机数一直是计算机科学的重要课题. 以从 [0, 1] 区间上取随机数为例，取出的单个数串应在 [0, 1] 上均匀分布（当数串足够长时），而当把每一对数看作平面单位正方形中点的坐标时，这数串应在 [0, 1] × [0, 1] 上均匀分布…，而在 n 维的正方体 [0, 1] ×…× [0, 1] 中每 n 元数组作为其中的点也应是均匀分布的. 图4.2所示的是由 TURBO PASCAL 的随机数生成器所产生的随机数在正方形内的点图，规则的条纹仍揭示了伪随机数的不均匀性. 这种点图的办法显得笨拙而永无尽头，不能令人满意.

有人定义一串数（比如0，1数串），如果无法设计比原来数

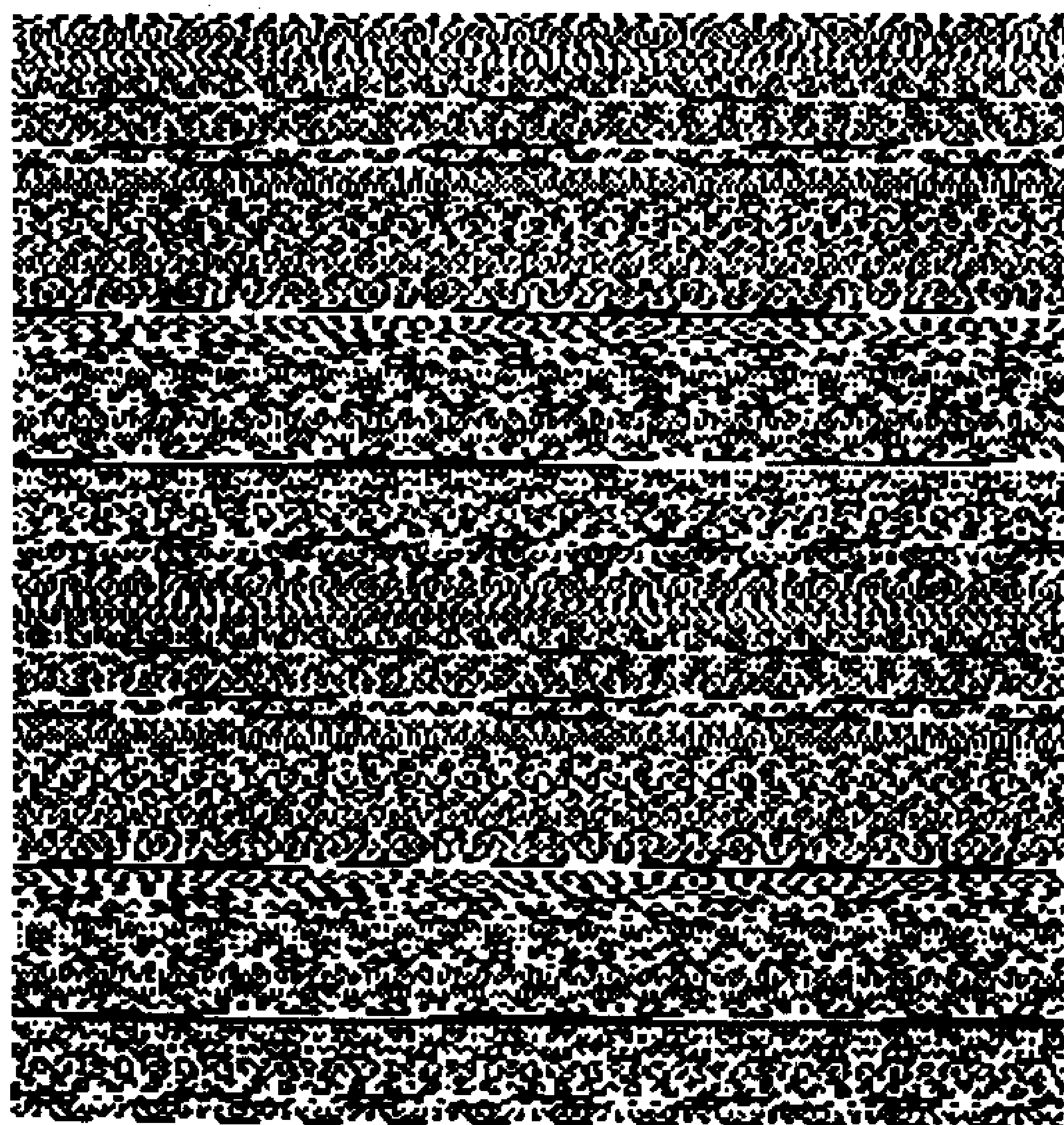


图4.2

串更短的二进位数码的算法来确定它就是随机的．比如01010000001010011011就显然比01010101010101010101有更好的随机性．而后者 可以由“01”的简单重复而生成．现实中的随机数也可由物理设备来产生：扔钱币和骰子，宇宙射线计数，查随机数表等．在IBM/PC机上常用的递归算法有

$$X_i = (aX_{i-1} + c), \qquad (\mathrm{mod}\ m)$$

其中$m=2^{31}-1$，$a=630360016$等，后来也有人设计素多项式的递归算法，1986年，Wolfram利用杨辉三角细胞自动机（cellular automata）

$$X_{n+1}(i) = X_n(i-1)\ XOR\ (X_n(i) OR X_n(i+1))$$

产生随机数，颇令人称道．近700年前中国古代的文化瑰宝在20

世纪的现代计算机科学中重放异彩.

依统计学的眼光看一串0、1数，若每个数 X_i 是独立同两点分布 $P(X_1=1)=P(X_1=0)=\frac{1}{2}$ 的，则应视 $X_1X_2X_3\cdots$ 为一串随机的0、1数串. 这种随机性恐怕仅存在于理论的笔端而永远无法在现实世界中实现的. 尽管如此，统计学还是设计了许多方法来检验经验观察数据的随机性，以从［0，1］中随机取数为例.

（1）χ^2拟合优度检验：把样本空间划分成 n 个子集、记 f_i 为样本落入第 i 个子集中的经验频数，p_i 为依理论确定的频数. 统计量

$$\chi^2=\sum_{i=1}^{n}\frac{(f_i-p_i)^2}{p_i}$$

则为用以检验随机性的统计量. 不计算统计量的分布，便无法对 χ^2值的大小作出评价. χ^2当然越小越表明 f_i 与 p_i 很接近. 若 X_1，X_2，…，X_n 为i.i.d $\sim N(0,1)$，则 $\sum_{i=1}^{n}X_i^2\sim\chi_n^2$，即自由度为 n 的 χ^2分布，一般统计教本上都有它的图和数表. 而我们定义的 χ^2统计量则服从自由度为 n-k-1的 χ^2分布记为 χ^2_{n-k-1}. 其中 k 为理论分布中有待估计的参数的数目. 对本例 $k=0$，当由实际观察数算出的 χ^2值大于临界值 C 时则有理由怀疑均匀性和随机性. 因而拒绝随机性的假定. C 值可由查表获得，其值的调整依赖于对犯错误判断概率大小的要求，也即取 C 使 $P(\chi^2>C)=\alpha$. 比如观察样本100个，把［0，1］区间分成10等分，则若 X_1，…，X_{100}是随机的，p_i 应为10. 记录每小隔中的 X_i 的数目就容易求出 χ^2的值.

（2）Kolmogrov-Smirnov 检验：基于样本 X_1，…，X_n 构造经验分布函数

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{(x_i \leqslant t)}.$$

由§2.2的讨论知统计量

$$D_n = \sup_{0\leqslant t\leqslant 1} \sqrt{n}\,|F_n(t) - t|.$$

当 n 很大时与布朗桥过程 $B(t)$ 的相应泛函

$$D = \sup_{0\leqslant t\leqslant 1} \sqrt{n}\,|B(t)|$$

的分布相同. 而后者的分布为

$$P\{D > d\} = 2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2k^2d^2}, \quad d > 0.$$

这样又可选取 d 来判断从实际观察的数据是否组成了与 $U(0,1)$ 的分布函数 $F(t)=t$ 很接近的 $F_n(t)$，量 D 衡量了两者的最大距离. D 过大则应否定随机性.

(3) 游程检验：将［0，1］中的数写成二进制小数，于是有了0，1组成的数串. 而对0，1组成的数串有一种游程检验方法，记数串中连续相同数码组成的一段为一个游程（run)，其长度为数码的个数. 比如000101100中有3个“0”游程，2个“1”游程，长度分别为3、1、2及1、2. 假定总共有 m 个1，n 个0，记 $m+n=N$. 并记 R 为游程总数，则当数串是随机的（0，1出现的概率相同，且前后各数码的出现相互独立），则有

$$P\{R=r\} = \begin{cases} \dfrac{2\dbinom{m-1}{k-1}\dbinom{n-1}{k-1}}{\dbinom{m+n}{m}}, & r = 2k; \\[2ex] \dfrac{\dbinom{m-1}{k-1}\dbinom{n-1}{k} + \dbinom{m-1}{k}\dbinom{n-1}{k-1}}{\dbinom{m+n}{m}}, & r = 2k+1. \end{cases}$$

更若假定当 $m\to\infty$，$n\to\infty$，$\frac{m}{n}\to\lambda$，统计学家 Wald 和 Wolfowitz 证明了在随机性的假定下，统计量

$$Z=\frac{R-1-\frac{2mn}{N}}{\sqrt{\frac{2mn(2mn-N)}{N^2(N-1)}}}$$

将依分布收敛到标准正态分布 $N(0,1)$．当 Z 过大或过小时均否定随机性．临界值可由正态分布的数值表查出．例如在由12个1和8个0组成的数串中出现6至14个游程的概率为0.95．我们在本节见到的第一个数串有12个游程不能否定其随机性．而第二个则依相应的统计量必能否定其随机性．

（4）常返时检验：记 X_1，…，X_n 为 $[0,1]$ 中的 n 个观察值．$S_i=X_i-X_{i-1}$，$i=1$，…，$n+1$，其中 $X_0=0$，$X_{n+1}=1$．记

$$S=\frac{1}{2}\sum_{i=1}^{n+1}\left|S_i-\frac{1}{n+1}\right|$$

为 Sherman 统计量，当 S 偏大时也可以否定 X_i 的均匀和随机性．

确定性模型所产生的序列也容易通过上述的某些检验鱼目混珠，使究竟何为随机的问题变得复杂．目前能为人们接受的观点似乎应该是：混沌与纯粹随机都有近似的表象，前者有确定的低维，简单的结构，而随机性则是高维，甚至无穷维的确定性系统．因而找寻混沌也就是要发现产生复杂随机现象的简单机制．这样统计学面临寻找潜结构的难题．

2．发现潜结构及时滞再造

假如一位朋友交给你由 $f(x)=4x(1-x)$ 对某一个初值经过10000次迭代而获得的数据，却并不告诉你这堆数据的由

来，请你进行统计分析．你能做的是首先点出序列的点图，并检验其随机性．粗心的分析者可能会得出结论说这是一堆随机数！是在［0，1］上均匀分布的．细心的人会对 X_{n+1} 关于 X_n 点图发现了 $X_{n+1} \doteq 4X_n(1-X_n)$ 的秘密．但如果原始数据依其落在 $\left[0, \frac{1}{2}\right)$ 和 $\left[\frac{1}{2}, 0\right]$ 而以01001…的 Σ_2 上的符号动力体系形式出现时．点图法失效了．所以统计学家面临的又是一个如何能有效地发现潜结构的问题．随数据表现形式的不同还有实验设计的问题．对高维的动力系统实验数据，若能通过提取了原系统主要信息的低维数据来进行处理就更为方便．方法之一是 Poincaré 截面．此方法十分有力，比如对 Lorenz 系统，取三维空间的平面 $\{(x,y,z)|x=y\}$ 为 Poincare 截面．基于原数据可以构作一个新的时间序列 $\{\cdots x(s), x(s+1)\cdots\}$. $x(s)$ 为原系统的轨道在第 s 次穿越截面时所处的位置．重要的是这个低维的离散时间序列保持原系统的若干重要的定性特征．事实上还存在一个迭代函数 g 使 $g(x(s))=x(s+1)$，g 又称为 Poincáre 映射．图4.3所描出的是由 Berliner 基于20000个数据（迭代数值解）所得到的与 Poincare 截面的400个交点而绘出的 g 的大致图形．这种方法也为寻找潜结构提供帮助．

在混沌数据分析中的另一个重要论题是时滞再造(time delay reconstruction)．对整个系统的多个变量的全方位的追踪考查在实践中几乎是很难实现的，人们只能对最有代表性的指标进行观测和记录，得到时间序列 $\{U_t, t\in T\}$，当 t 足够大，观察足够多能否以时间为代价，获得系统演化的空间规律呢？初看这几乎不可能．但是由 Packard（1980）和 Takens（1981）等人发展了一种从一个单变量时间序列出发再造出系统吸引子（的一个同构映象）的方法称为时滞再造．这是一种有趣的时空

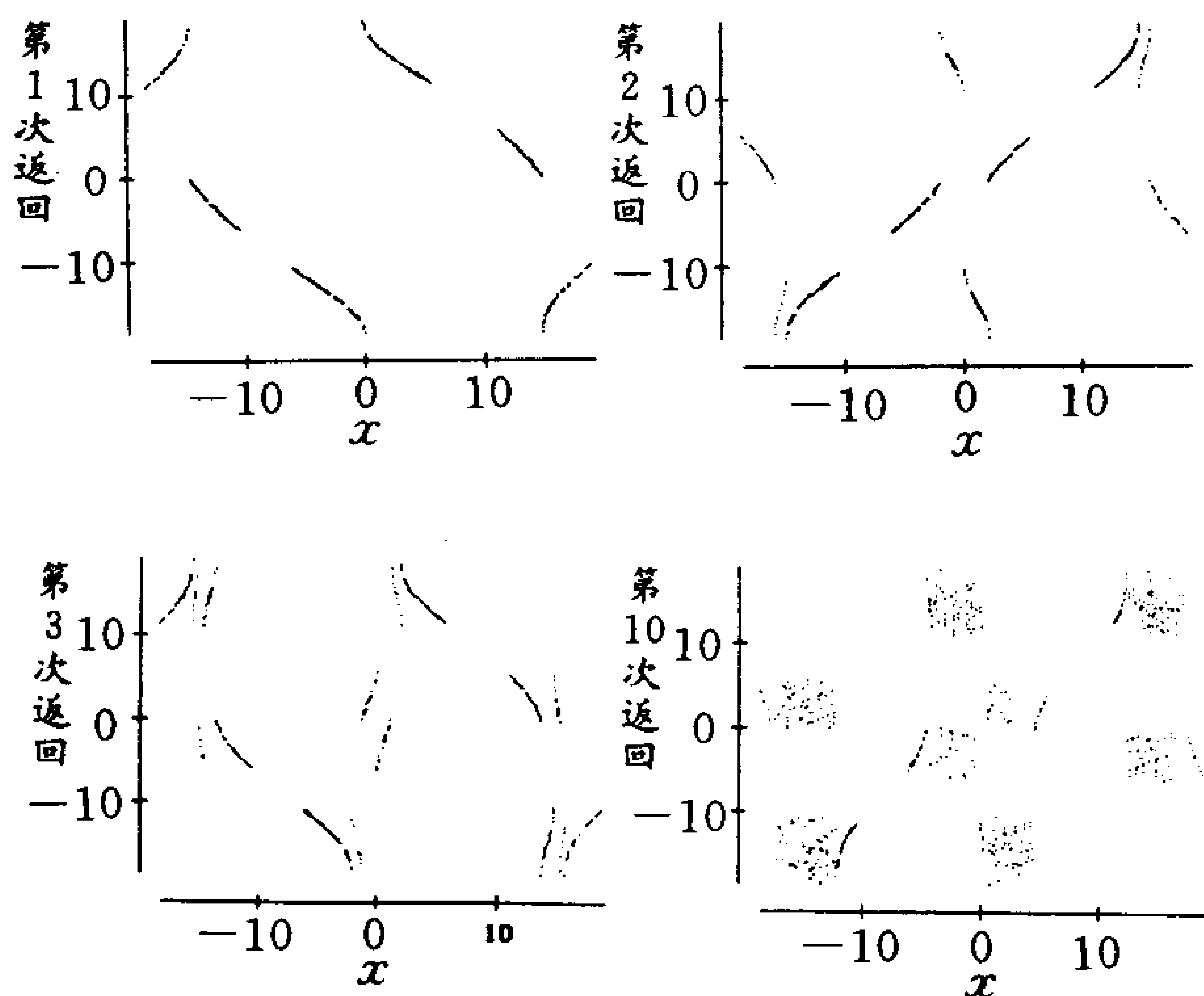


图4.3

转换模式.

令 $T:\mathscr{X}\rightarrow\mathscr{X}$ 为两次连续可微的且具有嵌在 d 维紧流形 M 的吸引子 A. $h:M\rightarrow R$ 为 d 维流形到实直线的映射（可以理解为某种投影）. 记 $U_k=h\,(T^k\,(X))$，$k=1，2，\cdots$，$X\in M$ 为随机的初始值. 则有

定义 对固定的整数 $\tau\geqslant2$，R^τ 中的 $\{V_k\}$，$V_k=(U_k,U_{k+1},\cdots,U_{k+\tau-1})$，$k=1,2,\cdots$，称原系统的嵌入维数为 τ 的时滞嵌入.

Takens、Mấne 等人证明了在一定条件下当 $\tau>2d$（$M\times M$ 的维数）时 V_k 的渐近性状是与 T^k（X）类似的. 具体讲他们证明了映射 Φ_τ：

$$x \to (h(x), h(T(x)), \cdots, h(T^{\tau-1}(x)))$$

是流形 M 到 R^{τ} 的一个嵌入.（所谓嵌入是带连续可微反函数的连续可微1—1映射）. 满足上述结果的最小 τ 记为 τ^*，则称为是系统的极小嵌入维数. 于是只要 τ 足够大原来难以观察的系统演化 T^k（X）可由嵌入向量 V_k 在微分同胚的意义下再造出来. 吸引子的维数也将是这一 Φ_τ 映射下的不变量，即有 $D_H(A) = D_H(\Phi_\tau(A))$. 而且若 μ 为 A 的遍历分布，则 $\mu\Phi_\tau^{-1}$ 也将是 Φ_τ（A）上的遍历分布. 而且它们有相同的维数分布.

例 Hénon 映射

$$H(x,y) = (1 + y - 1.4x^2, 0.3x),$$

取 $h(x,y) = x, T = H$，也即仅观察第一个坐标. 这时 $\tau^* = 2$，$V_k = (X_k, X_{k+1}), \Phi_\tau = \Phi_2 = (x, 1 - y - 1.4x^2)$ 为双向满足 Lipschtz 条件（比如导数有界就能使映射满足 Lipschtz 条件）的同胚. 图4.4绘出由 (X_k, X_{k+1})，$k=1,2,\cdots,2999$ 得出的从时滞再造中得到的 Hénon 吸引子. 读者不妨将它与图3.8对照.

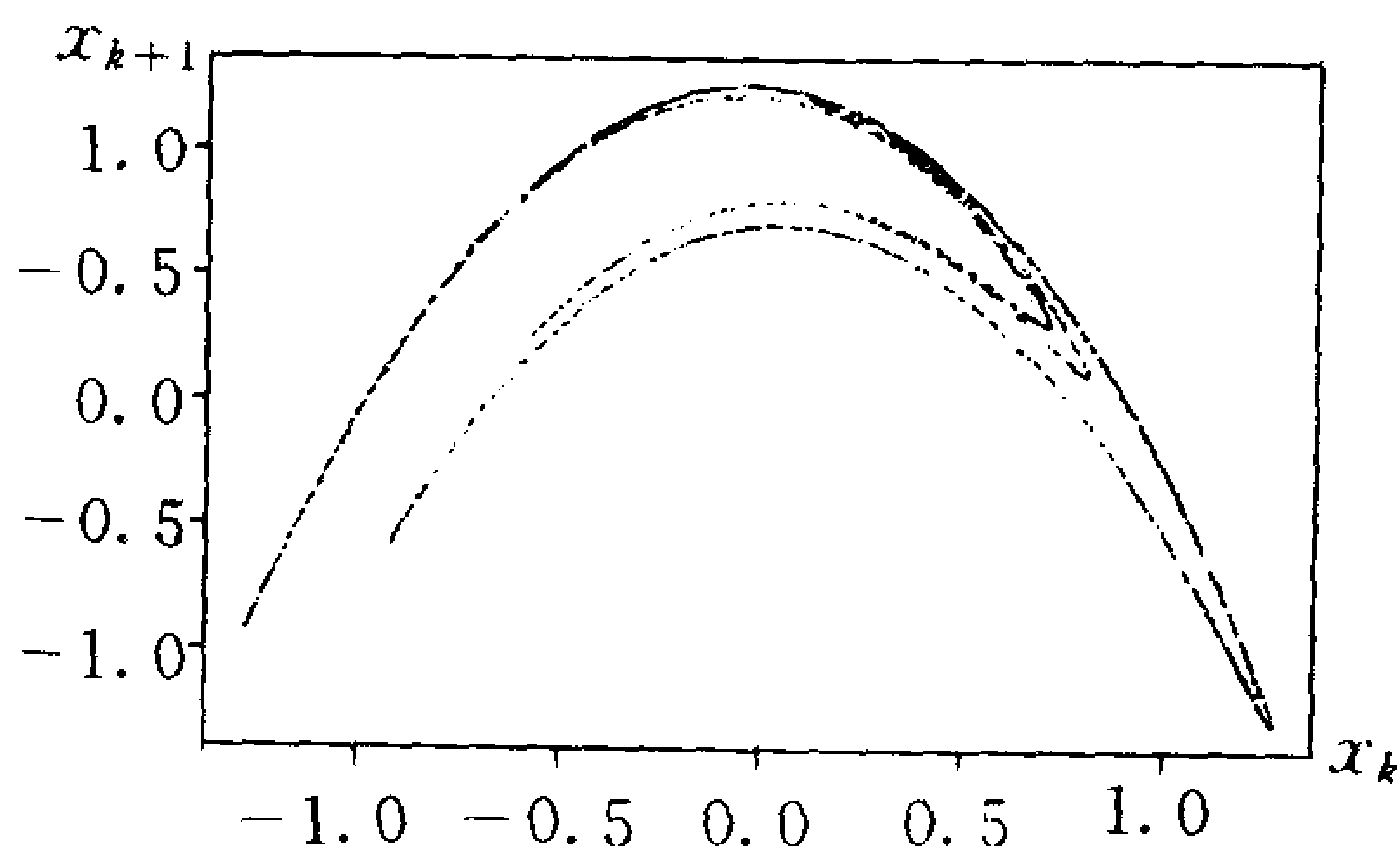


图4.4

对连续时间动力系统，可利用时滞参数 $S>0$ 来构作嵌入向量

$$V_t=(U_t,U_{t+S},\cdots,U_{t+(\tau-1)S}).$$

实际构作中要注意 S 的选取. S 选得过小，V_t 将挤在 R^τ 中的对角线上，无法辨识. S 选得过大，U_t 与 U_{t+S} 就会互不相关. 最佳时滞参数 S 的选择也是一个统计学问题，而更为重要的是最小嵌入维数 τ^* 的估计，它是搜索潜在结构的基础. 从某种意义上讲，高维的系统更接近于随机现象，而产生混沌的确定性系统则是低维的. 所以最小嵌入维数将是体现两者区别的重要特征. 从以下论证中也可以更清楚地了解这一点.

记 $\{U_k\}$ 为某一概率空间上的平稳时间序列. 记 $U_j,U_{j+1},\cdots,U_{j+(n-1)}$ 的联合分布为 P_n，$\|x-y\|=\max\limits_{1\leqslant i\leqslant n}|x_i-y_i|$ 为 R^n 中的距离. 则有

引理 若对 $n=1,2\cdots$，P_n 的相关维数 ν_n 存在则 $\nu_n\leqslant\nu_{n+1}$，$\nu_\infty=\lim\limits_{n\to\alpha}\nu_n$ 称为 $\{U_k\}$ 的相关维数.

证明 记 $C_n(\varepsilon)=P_n\times P_n(\|x-y\|\leqslant\varepsilon)$，则 $C_n(\varepsilon)=P_n\times P_n(\{x_1,\cdots,x_n;y_1,\cdots,y_n\}|\max\limits_{1\leqslant i\leqslant n}|x_i-y_i|\leqslant\varepsilon)$. 于是有

$$\begin{aligned}C_n(\varepsilon)&=P_{n+1}\times P_{n+1}(\{x_1,\cdots,x_{n+1};y_1\cdots y_{n+1}\}|\\&\qquad\max_{1\leqslant i\leqslant n}|x_i-y_i|\leqslant\varepsilon)\\&\geqslant P_{n+1}\times P_{n+1}(\{x_1,\cdots,x_{n+1};y_1,\cdots,y_{n+1})\}|\\&\qquad\max_{1\leqslant i\leqslant n+1}|x_i-y_i|\leqslant\varepsilon)\\&=C_{n+1}(\varepsilon).\end{aligned}$$

由 ν_n 之定义得证.

基于这一引理又有

定理4.1.1 (i) 若 $\{U_k\}$ 为 i.i.d 的时间序列，则 $\nu_n=n\nu_1$，若 $\nu_1>0$，则 $\nu_\infty=\infty$；(ii) 若 $\{U_k\}$ 为平稳时间序列，且存在 n_0 及满足 Lipschitz 条件的函数 f：$R^d\to R$ 使当 $k>n_0$时，$U_k=f(U_{k-n_0},\cdots,U_{k-1})$，则对 $n\geqslant n_0$有 $\nu_n=\nu_{n_0}$ (ν_{n_0}即为 P_{n_0}之相关维数).

证明 (i) 由独立性 $P_n=P_1\times\cdots\times P_1$，则 $C_n(\varepsilon)=[C_1(\varepsilon)]^n$，由定义 $\nu_n=n\nu_1$. 由引理 $\nu_{n+1}\geqslant\nu_n$，又由 $f:R^{n_0}\to R$ 满足 Lipschitz 条件知存在 $K\geqslant1$使

$$|f(x_1,\cdots,x_{n_0})-f(y_1,\cdots,y_{n_0})|\leqslant K\max_{1\leqslant i\leqslant n_0}|x_i-y_i|.$$

当 $n\geqslant n_0$，

$$\begin{aligned}C_{n+1}(\varepsilon)&=P_{n\times1}\times P_{n+1}(\{x_1,\cdots,x_{n+1};y_1,\cdots,y_{n+1}\}\mid\max_{1\leqslant i\leqslant n+1}|x_i-y_i|\leqslant\varepsilon)\\&=P_n\times P_n(\{x_1,\cdots,x_n;y_1,\cdots,y_n\}\mid\max_{1\leqslant i\leqslant n}|x_i-y_i|\leqslant\varepsilon\text{ 且}\\&\qquad|f(x_{n+1-n_0},\cdots,x_n)-f(y_{n+1-n_0},\cdots,y_n)|\leqslant\varepsilon)\\&\geqslant P_n\times P_n\left(\{x_1,\cdots,x_n;y_1,\cdots,y_n\}\mid\max_{1\leqslant i\leqslant n}|x_i-y_i|\leqslant\frac{\varepsilon}{K}\right)\\&=C_n\left(\frac{\varepsilon}{K}\right).\end{aligned}$$

因此 $$\nu_{n+1}=\lim_{\varepsilon\to0}\frac{\log C_{n+1}(\varepsilon)}{\log\varepsilon}\geqslant\lim_{\varepsilon\to0}\frac{\log C_n\left(\frac{\varepsilon}{K}\right)}{\log\left(\frac{\varepsilon}{K}\right)}\,\frac{\log\left(\frac{\varepsilon}{K}\right)}{\log\varepsilon}=\nu_n.$$

故有当 $n\geqslant n_0$时，$\nu_{n+1}=\nu_n=\nu_{n_0}$. □

结果(i)表明独立序列表现出很大的随机性，而结论(ii)所

显示的是过程或序列的全部变化性和随机性都局限在前 n_0 个变量，所有其他变量均为 U_1，…，U_{n_0} 的函数. 维数估计于是成为区分确定性混沌与随机序列的重要工具. 嵌入维数的估计与线性 **ARIMA** 模型的模型选取及定阶极为类似，目前流行的定阶准则为 AIC，赤池信息论准则，BIC 基于 Bayes 观点考虑的准则. 我国学者赵林城博士则发展了一种独创的更为精确的判别准则，不妨称之为 ZIC 准则，其方法的要旨是利用收敛速度的差别来判别. 非线性模型的定阶及最小嵌入维数的估计是混沌学统计研究的重要课题，有待引起关注.

§4.2 分形维数的估计

在维数定义中，考察的是某个量 $H(\varepsilon)$(比如 $C_n(\varepsilon)$，$N_n(\varepsilon)$ 等)，当尺度 $\varepsilon \to 0$ 时若极限存在就是相应的维数 d，也即

$$\frac{\log H(\varepsilon)}{\log \varepsilon} = d + h(\varepsilon).$$

当 $\varepsilon \to 0$ 时，$\lim\limits_{\varepsilon \to 0} h(\varepsilon) = 0$，于是 $H(\varepsilon) = k(\varepsilon)\varepsilon^d$，其中 $k(\varepsilon) = \varepsilon^{h(\varepsilon)}$. 当 $\varepsilon < \varepsilon_0$ 时，若有 $k(\varepsilon) = k$，则 $H(\varepsilon)$ 服从指数律 $k\varepsilon^d$. 但 $k(\varepsilon)$ 的极限可能不存在. 这将导致对 d 的估计的不相容性，这一点需在实际数据的估计中引起注意. 这种现象被 Mandelbort 称为空洞，也已有人研究空洞的度量. 在维数估计中要区分待估计的是空间的几何分形集的维数还是以分形集为支撑的某个概率分布的维数. 在以 μ 为不变分布的平稳时间序列中碰到的维数估计问题属于后者. 常用的估计方法首先是对 $H(\varepsilon)$ 作出某种估计 $\hat{H}(\varepsilon)$，然后取一串 $0 < \varepsilon_1 < \varepsilon_2 < \cdots < \varepsilon_m$ 对平面的点 $(\log \varepsilon_i, \log \hat{H}(\varepsilon_i), i = 1, \cdots, m)$ 配最小二乘直线(即求直线与各点距离

最小)，该线的斜率即是 d 的估计 $\hat{d}$. 与通常回归中求最小二乘估计的不同之点是 $\hat{H}(\varepsilon_i)$ 对不同 i 是相关的，估计的误差分析则要依赖 $\log\hat{H}(\varepsilon_i)$ 的协方差矩阵作必要的修正. 关于 $\hat{H}$ 的估计又分为计盒，样本相关积分、近邻方法、极大似然估计等多种方法. 这里略作介绍：

(1) 计盒方法相对简单，将待估的集合 E 作适当的刻度变换，使它能包含在一个单位立方体中（N 维）. 以边长为 $\frac{1}{r^n}$ 的，由 r^{nN} 个小盒组成的网格来覆盖 E，记 $N_n(E)$ 为盖住 E 的小格子的数目. 将 $\log N_n(E)$ 对 $n\log r(n=j,j+1,\cdots,j+S)$ 配回归线，通常对数据总量 t 要求很大，否则 n 次之后网格中就没有 E 中的点了.

在估计概率分布的维数时应先估计分布 μ. 记水平为 n 的网格中第 j 个小盒为 $C_{n,j}$，记落入 $C_{n,j}$ 中的集合 E 的点数为 $N_{n,j}(E)$，并以经验频数 $N_{n,j}(E)/t$ 为 $\mu(C_{n,j})$ 的估计. 由信息维数的定义，应先估计量

$$H_q(r^{-n})=\left(\sum_{j=1}^{r^{nN}}\mu(C_{n,j})^q\mu(C_{n,j})\right)^{\frac{1}{q}},$$

其经验估计为

$$\hat{H}_q(r^{-n})=\left(\sum_{j=1}^{r^{nN}}\left(\frac{N_{n,j}(E)}{t}\right)^q\left(\frac{N_{n,j}(E)}{t}\right)\right)^{\frac{1}{q}},$$

而

$$\hat{D}_{R,q}=\lim_{n\to\infty}\frac{\log\hat{H}q(r^{-n})}{\log(r^{-n})}$$

为 $D_{R,q}$ 的估计. 这种方法失之过粗，而且 $\hat{D}_{R,q}$ 趋于 $D_{R,q}$ 的速度很慢，易受空洞的影响而使估计不准确. 对计盒维数的改进方案已有不少人研究.

(2) 相关维数的估计依赖于相关积分 $C_\mu(\varepsilon)$，即 $\mu\times\mu$

$(\|x-y\| \leqslant \varepsilon)$, x, $y \in \boldsymbol{R}^D$, 的估计. 与计盒方法相比有更快的收敛速度，加之有 Grassberger-Procaccia 的著名算法，使其广为应用工作者所接受，这是当前最实用的维数估计技术. 尽管所估计的是分布 μ 的维数 ν，但它是吸引子的几何分形维数 很好的下界（回想 $D_B \geqslant D_H \geqslant D_I \geqslant D_C$）. $C_\mu(\varepsilon)$ 的最自然的估计是样本相关积分. 记 V_1，…，V_n 为 R^D 中的观察向量，任取一对向量将它们的距离与 ε 相比，小者记1，大者记0. 把所有$\binom{m}{2}$对 (V_i, V_j) 之中小于 ε 的对子数的比例作为 $C_\mu(\varepsilon)$ 的估计. 写下来即

$$C_n(\varepsilon) = \binom{n}{2}^{-1} \sum_{i \neq j} 1_{(\|V_i - V_j\| \leqslant \varepsilon)}.$$

这是典型的以$1_{(\|V_i - V_j\| \leqslant \varepsilon)}$为核的 U-统计量. 有关 U-统计量的介绍可参看陈希孺著的《非参数统计》，上海科技，1989年版，估计量 $C_n(\varepsilon)$ 将以各种收敛方式当 $n \to \infty$ 时收敛到 $C_\mu(\varepsilon)$. 在 V_i 相互独立时结论是经典的. 对 V_i 是混合序列的情形 Denker 和 Keller 等人作了研究. 确而言之有

定理4.2.1 若 V_1，V_2，…，为 R^D 中的平稳序列，边缘分布为 μ，且当 $n \to \infty$时对某个 $\delta > 0$，混合系数 $\beta(n) = O(n^{-(1+\delta)})$（混合系数刻画了当 V_i, V_j 在 $|i-j| > n$ 时的互相关联程度），对固定的$0 < \varepsilon_1 < \cdots < \varepsilon_k$，记

$$C_n(\varepsilon) = (C_n(\varepsilon_1), \cdots, C_n(\varepsilon_k))^\tau,$$

$$C_\mu(\varepsilon) = (C_\mu(\varepsilon_1), \cdots, C_\mu(\varepsilon_k))^\tau,$$

（此处，τ 代表向量转置），则几乎处处有

$$C_n(\varepsilon) \to C_\mu(\varepsilon),$$

$$\log C_n(\varepsilon) \to \log C_\mu(\varepsilon).$$

且存在非负矩阵 U、V 使

$$\sqrt{n}\,(C_n(\varepsilon)-C_\mu(\varepsilon))\xrightarrow{\text{依分布}}N_k(0,U),$$

及
$$\sqrt{n}\,(\log C_n(\varepsilon)-\log C_\mu(\varepsilon))\xrightarrow{\text{依分布}}N_k(0,V),$$

其中 $N_k(o,\Sigma)$ 为 k 维正态分布, $U=[u_{ij}]$, $V=[v_{ij}]$, $v_{ij}=\dfrac{u_{ij}}{C_\mu(\varepsilon_i)C_\mu(\varepsilon_j)}$.

当 V_i 独立时

$$u_{ij}=4\mathrm{Cov}(\mu(B(V,\varepsilon_i)),\mu(B(V,\varepsilon_j)))\stackrel{\triangle}{=\!=}4\tau(\varepsilon_i,\varepsilon_j),$$

当 V_i 混合时

$$u_{ij}=4\tau(\varepsilon_i,\varepsilon_j)+8\sum_{m=1}^{\infty}k(m,\varepsilon_i,\varepsilon_j),$$

其中 $k(m,\varepsilon_i,\varepsilon_j)=\mathrm{Cov}(\mu(B(V_1,\varepsilon_i)),\mu(B(V_{m+1},\varepsilon_j)))$.

基于上述定理不仅能获得估计量 $C_n(\varepsilon)$,而且对估计量的优良性及与 $C_\mu(\varepsilon)$ 的误差有了较确切的了解. 估了 $C_\mu(\varepsilon)$,则当 $C_\mu(\varepsilon)\sim k\varepsilon^{\nu_n}$ 时 ν_n 为所求的相关维数.

目前物理学界在实际中应用的 Grassberger-Procaccia 算法 (GP 算法) 是将时滞嵌入向量与相关维数的估计相结合. 大意为: 从观察序列 $\{U_k\}_k$ 出发对确定的 D 构作嵌入向量

$$V_k(D)=(U_k,U_{k+1},\cdots,U_{k+D-1}),$$

$k=1, 2, \cdots, n-D+1$. 计算 $V_k(D)$ 的样本相关积分 $C_{n-D+1}(\varepsilon)$, 并由此估计出 $\hat{\nu}_D$. 对 $D=1, 2, \cdots$, 重复上述步骤, 直至 $\hat{\nu}_D$ 不随 D 的增加而增加时停止. 据定理4. 1. 1取 $\lim\limits_{D\to\infty}\hat{\nu}_D$ 为原分布 μ 的相关维数.

(3) 近邻方法.

在计盒方法及相关积分方法中半径 ε 是固定的, 而近邻方

法的要旨是半径是随机的，以逐点维数 $\alpha_\mu(x) = \lim_{\varepsilon\to 0}\dfrac{\log\mu(B(x,\varepsilon))}{\log\varepsilon}$ 为例，以 x 为中心固定 ε 考查小球 $B(x,\varepsilon)$ 中有多少点与依次考查与 x 最接近的点与 x 的距离有多远，所得到的关于这些点疏密的信息大致是一样的. 从一串 i. i. d 的服从分布 μ 的 $V_1, \cdots, V_n$ 出发记 $\delta_j(x) = \|V_j - x\|, j = 1,2,\cdots,n$，将 $\delta_j(x)$ 排序为 $\delta_{(j)}(x)$，使

$$\delta_{(1)}(x) < \cdots < \delta_{(n)}(x).$$

对选定 k 将 $\log\delta_{(j)}(x)$ 关于 $\log\left(\dfrac{j}{n}\right)$ 作最小二乘配线，$j=1,\cdots,k$，此斜率的倒数即为 $\alpha_\mu(x)$ 之估计，也即

定理 4.2.2 若 $\alpha_\mu(x)$ 存在，则几乎处处有

$$\lim_{n\to\alpha}\frac{\log\delta_{(j)}(x)}{\log(j/n)} = \frac{1}{\alpha_\mu(x)}.$$

有关正则化之后该统计量的渐近分布也已有了深入的研究，因其次序性质其分布与极值分布有关. 因 $\sigma_j(x)$ 取代尺度 ε 的地位，所以自然所得到的是 Hölder 指数估计的倒数.

（4）Lyapunov 指数估计.

在混沌的研究中，最大的 Lyapunov 指数 λ_1是否大于零是十分重要的特征指标. 有关 λ_1的估计因而也特别令人感兴趣. 现有一方法是由 Guckenheimer（1982）提出，Wolf（1985）等人发展的直接法. 仍记 $\{x_t\}$ 为时间序列，考虑到观察可能出现误差，用来估计 λ_1的模型为

$$X_t = f(X_{t-1}) + \sigma\varepsilon_t.$$

在 $\{X_t\}$ 中取 (t_1, t_2) 使$\|X_{t_1}-X_{t_2}\|$充分小，比如 ε. 当经过一段时间 τ 之后，$\|X_{t_1+\tau}-X_{t_2+\tau}\|$超过某一预定的界限 M 时，再选

取 X_{t_3} 与 $X_{t_1+\tau}$ 充分接近，而且与 $X_{t_1+\tau}$ 与 $X_{t_2+\tau}$ 的连线也尽量接近，对 $\|X_{t_1+\tau+\sigma}-X_{t_3+\sigma}\|$ 进行观察，当它超过 M 再同样找 X_{t_4}. 不断继续进行之后，求出整个数据集 $\{X_t\}$ 上的平均发散速度，做为 λ_1 的估计. 方法之二是所谓"间接法". 先对映射 f 的 Jacobian 进行估计，并代入 $T^t=J_{t-1}J_{t-2}\cdots J$，得出 T^t 之估计，记 ν_1 为 $T^{t^*}T^t$ 的最大特征根. $\hat{\lambda}_1=\frac{1}{2t}\log\nu_1$ 即为 λ_1 之估计，而对 f 的 Jacobian 的估计又依赖于对 f 本身的估计. 后者本质上是一个非参数回归问题. 非线性回归、参数与半参数回归的模型在统计中内容丰富. 中国科大陈希孺、赵林城教授在这些方面作了大量工作，这些成果的应用将有助于 f 的估计. 而整个 Lyapunov 谱的估计则是个有待充分研究的课题，相应的 Lyapunov 维数 D_L 的估计也更具挑战性.

有人讲过统计不是科学而是艺术，统计方法很象雕刻家手中的刀. 艺术家能借助它刻出精美的艺术品，而初学者却难以运用自如. 统计方法常显得零散，不如传统数学那么漂亮，但经验的积累，方法的更新和灵活的调度是这门学科的魅力所在. 统计学科所展示给社会的是精巧的方法和丰富的统计思想，它所提供的模型将使各门学科受益.

§4.3 非线性时间序列模型

非线性时间序列是统计学者与动力系统研究者两群体研究领域的结合，互相影响日趋增加. 在动力系统 $X_t=f(X_{t-1})$ 中以某种方式加入误差项 ε_t 之后，被观察到的系统演化就成为典型的非线性时间序列. 而对混沌科学中的数据分析而言有参数

和非参数模型两大类. 例如在模型 $X_t = \mu X_{t-1}(1 - X_{t-1}) + \varepsilon_{t-1}$ 中,函数形式已知,但 μ 可能未知,对 μ 及 ε_t 的估计属参数模型. 若 $f(x)$ 的形式本身未知,则属非参数模型类. 人们力图建立恰当的模型来拟合系统 $X_t = f(X_{t-1})$. 与通常线性的时间序列模型相平行已有了一整套关于非线性时间序列的建模，稳定性条件分析，可逆性研究，模型选择、诊断以及更加困难的预测和控制的理论及方法. 相应的统计软件包也已付诸实用. 与线性模型不同的则还有非线性的检验问题，但从总体看非线性时间序列分析由于问题本身的复杂性，是有待深入的研究领域. 借助非线性时间序列分析的研究来认识混沌，走出混沌是个值得大力推进的方向.

1. 非线性自回归模型(NLAR 模型)

自回归模型 (AR 模型) $\phi(B)X_t = \varepsilon_t$ 的最自然的推广是非线性自回归模型 NLAR：

$$X_t = g(X_{t-1}, X_{t-2}, \cdots, X_{t-k}, \varepsilon_t), \quad t \in \boldsymbol{Z}.$$

ε_t i.i.d 且与 X_S，$s<t$ 独立. 经过变换

$$\boldsymbol{Y}_t = (Y_t^{(1)}, \cdots, Y_t^{(k)})^\tau = (X_t, X_{t-1}, \cdots, X_{t-k+1})^\tau,$$

$$\eta_t = (\eta_t^{(1)}, \cdots, \eta_t^{(k)})^\tau = (\varepsilon_t, 0, \cdots, 0)^\tau.$$

记 $f: R^{2k} \to R^k$ 为

$$Y_t^{(1)} = g(Y_{t-1}^{(1)}, \cdots, Y_{t-1}^{(k)}, \eta_t^{(1)}),$$

$$Y_t^{(2)} = Y_{t-1}^{(1)},$$

$$\cdots\cdots$$

$$Y_t^{(k)} = Y_{t-1}^{(k-1)},$$

则有 $\boldsymbol{Y}_t = \boldsymbol{f}(\boldsymbol{Y}_{t-1}, \eta_t)$,且 η_t 本身 i.i.d 与 $Y_s, s < t$ 独立. 当 $\eta_t = 0$ 则回到了动力系统 $\boldsymbol{Y}_t = \boldsymbol{f}(\boldsymbol{Y}_{t-1})$.

例 以已熟知的 Henon 系统为例，在系统方程组中将

$$\begin{cases} x_{k+1} = y_k + 1 - a x_k^2, \\ y_{k-1} = b x_k \end{cases}$$

的第二式代入第一式即有

$$x_{k+1} = - a x_k^2 + b x_{k-1} + 1.$$

即为 $\varepsilon_t = 0$时的 NLAR 模型，a、b 成了回归系数.

例 确定性的 Logistic 映射在时间倒置的作用下可变为带噪声乘数的 NLAR 模型.

$$X_{t-1} = 4X_t(1 - X_t), \quad t = 0, -1, \cdots, X_0 \in (0,1).$$

假定 X_0～反正弦分布，从 X_{t-1}到 X_t 的映射分为两支 $X_t = \frac{1}{2}\left[1 \pm \sqrt{(1 - X_{t-1})}\right]$，于是

$$X_t = \frac{1}{2}\left[1 + \varepsilon_t (1 - X_{t-1})^{\frac{1}{2}}\right], \quad t = 1, 2, \cdots$$

其中 $\varepsilon_t = \begin{cases} 1, & 以概率\frac{1}{2}, \\ -1, & 以概率\frac{1}{2} \end{cases}$ 确定的 Logistic 迭代与 NLAR 模型成为同一系统的不同表现形式.

由于模型 $X_t = f(X_{t-1}, e_t)$ 与离散时间的 Markov 链有天然的联系，有关 Markov 链的遍历性定理及有关连续状态 Markov 链的极限定理是在动力系统的研究中极为有用的. 这联系可表为

定理4.3.1 记 $\{X_t\}$ 为平稳实值的 Markov 链，有条件分布 $P\{X_t \leqslant y | X_t = x\} = F_x(y)$，则它存在随机变量 e_t 是关于 $\sigma(X_s, s \leqslant t)$ 可测，但与$\{X_s, s \leqslant t - 1\}$ 独立的充分必要条件是

$F_x(y)$ 与 $\{X_t\}$ 的不变测度等价.

证明 仅介绍思路：记 $F_x(y) = P\{X_{t+1} \leqslant y | X_t = x\}$，则 F_x^{-1} 为 F_x 的条件分位数

$$F_x^{-1}(u) = \inf\{y | F_x(y) > u\}, 0 < u < 1.$$

记 $U \sim U[0, 1]$，则 $F_x^{-1}(U) \sim F_x$. 若 $X_t = x$ 与 U 独立，则 $X_{t+1} = F_x^{-1}(U)$ 正有所需的分布. 假定 $U_1, U_2, \cdots$ 为 i. i. d $U[0, 1]$，定义 $f: R \times [0,1] \to R$，记 $f(x,u) = F_x^{-1}(u)$，则当 X_0 选取适当的初始分布 $X_{t+1} = f(X_t, U_{t+1})$ 给出的正是Markov 链 $\{X_t\}$ 的实现. □

2. 阈限模型，Threshhold 模型（TAR 模型）

用逐段线性函数来近似处理非线性函数的思想由来已久，而现实数据中也不乏这类模型，在昆虫的群体生物模型中有这种例子：

$$X_t = 0.8X_{t-1} + f(X_{t-15}).$$

产生这一模型的机制是虫卵要经过15天才能长为成虫，而昆虫群体的繁衍又依赖于食物的供给，但食物却是有限的. 当虫群过大食物的短缺使增长减缓，最后因食物耗尽而使增长为0. 0.8则表示成虫每天有20%会死去，于是

$$f(x) = \begin{cases} 10x, & \text{当 } x \leqslant 171, \\ 1795 - 0.503x, & \text{当 } 171 < x \leqslant 3569, \\ 0, & x > 3569. \end{cases}$$

f 是分段线性的. 阈限自回归模型的一类 TAR(P)，形如

$$X_t = \sum_{j=1}^{p} \varphi_j^{(i)} X_{t-j} + \varepsilon_t^{(i)}, \qquad (X_{t-1}, \cdots, X_{t-p}) \in R^{(i)}.$$

其中 $R^{(i)}$ ($i = 1, \cdots, k$) 为 R^p 之一个分划. $\varepsilon_t^{(i)}$ i. i. d 零均值，方差 σ_i^2 之误差白噪声序列.

在更一般的模型中 i 也可以是随机变量 J 的取值，这时有

$$\boldsymbol{X}_t = \boldsymbol{B}^{(J_t)}\boldsymbol{X}_t + \boldsymbol{A}^{(J_t)}\boldsymbol{X}_{t-1} + \boldsymbol{H}^{(J_t)}\varepsilon_t + \boldsymbol{C}^{(J_t)}.$$

对固定的 $J_t=j$，$\boldsymbol{A}^{(j)}$、$\boldsymbol{H}^{(j)}$为非随机的矩阵，而$\boldsymbol{C}^{(j)}$则为向量，考虑一个特殊情形：

例 $\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \boldsymbol{A}^{(J)}\begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \boldsymbol{G}^{(J)}$，取 $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$，$J$ 为四点分布

$J=$	1	2	3	4
P_j	0.01	0.85	0.07	0.07

$$A^{(1)} = \begin{pmatrix} 0 & 0 \\ 0 & 0.125 \end{pmatrix}, \qquad A^{(2)} = \begin{pmatrix} 0.85 & 0.04 \\ -0.04 & 0.85 \end{pmatrix},$$

$$A^{(3)} = \begin{pmatrix} 0.20 & -0.26 \\ 0.26 & 0.22 \end{pmatrix}, \qquad A^{(4)} = \begin{pmatrix} -0.15 & 0.28 \\ 0.26 & 0.24 \end{pmatrix},$$

$$C^{(1)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, C^{(2)} = \begin{pmatrix} 0 \\ 1.6 \end{pmatrix}, C^{(3)} = \begin{pmatrix} 0 \\ 0.8 \end{pmatrix}, C^{(4)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

读者能相象出这个模型代表什么吗?是如图4.5的分形集，称作羊齿草或 Barnsley Fern. 这也是不带可加随机误差项的 TAR 模型，随机部分在指标 J 上. 图4.6则显示了从 Koch 曲线在一定机制下变换成羊齿草的过程.

3. 双线性模型(BL 模型)

双线性模型植根于微分动力系统中的双线性系统. 很典型的例子是在生物群体模型中的

$$\frac{dx}{dt} = u(t)x(t).$$



图4.5

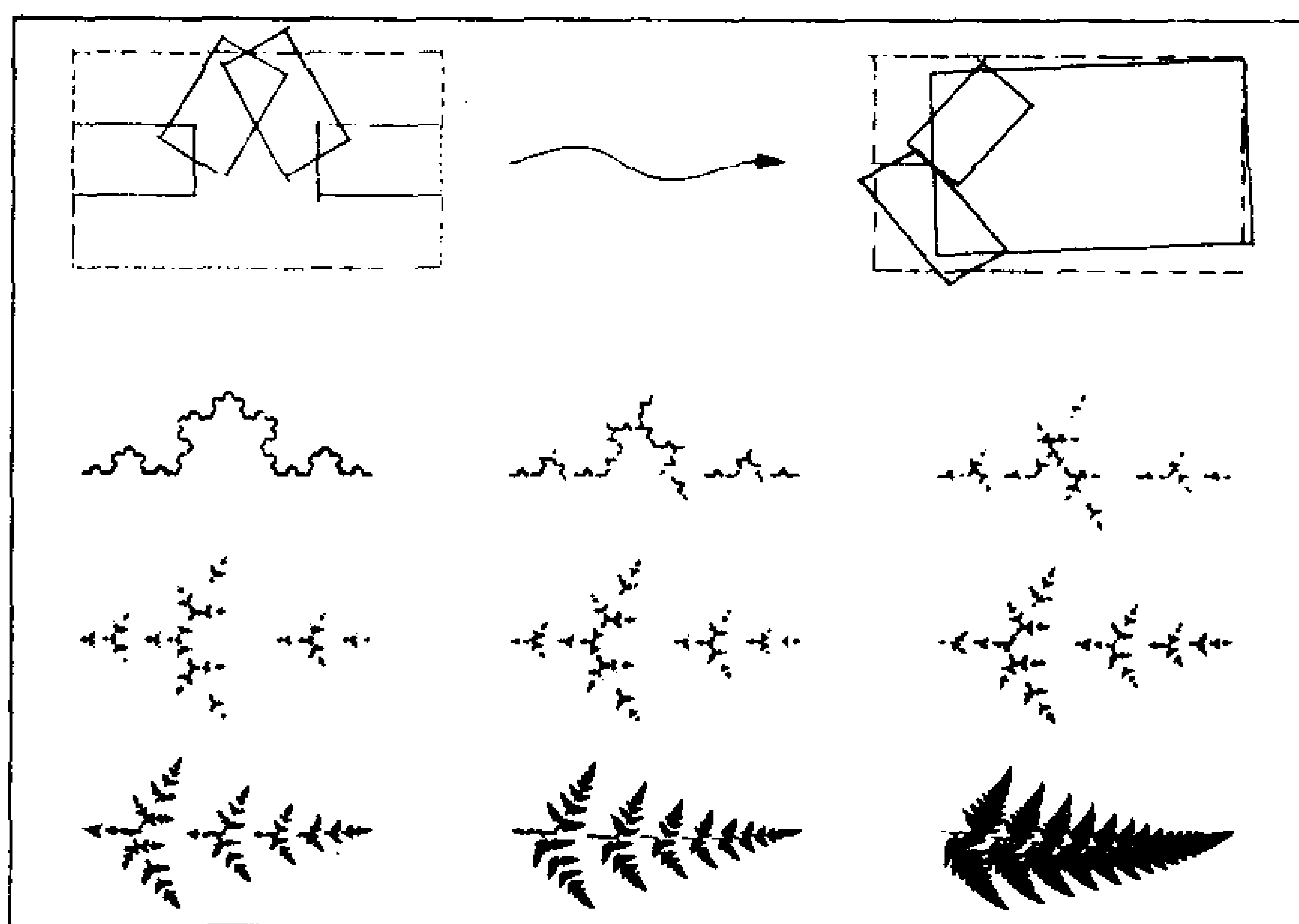


图4.6

其中 $x(t)$ 代表在时刻 t 群体的大小，$u(t)$ 是控制变量，它是出生率减去死亡率. $u(t)$ 又往往与 $x(t)$ 有关，其最简单的情况是

$u(t)=a-bx(t)$,相当于一种反馈作用. 于是回到了很熟悉的 Logistic 模型的微分形式. 双线性系统与有限 Volterra 级数($x(t)$ 的泛函的级数展开)

$$f(t;x(s)\ s\leqslant t)$$
$$=\sum_n\int\cdots\int k(t;t_1,\cdots,t_n)x(t_1)\cdots x(t_n)dt_1\cdots dt_n$$

有关. 当 k 可分解时可以证明系统必有双线性表示. 而非线性时间序列的 BL 模型形如

$$X_t=\sum_{j=1}^{P}\varphi_j X_{t-j}+\sum_{j=0}^{q}\theta_j\varepsilon_{t-j}+\sum_{j=1}^{P}\sum_{k=1}^{Q}b_{jk}X_{t-j}\varepsilon_{t-k}+\alpha.$$

其中 φ_j、θ_j、b_{jk}、α 为实数, ε_t 为 i. i. d 的误差序列. 固定 ε_s, $s<t$, X_t 关于 X_s, $s<t$ 线性. 固定 X_s, $s<t$, ε_t 关于 ε_s, $s<t$ 也是线性的. 双线性因此得名.

例 对时间倒置的帐篷变换 $X_{t-1}=2X_t$ (mod 1), $t=0$, -1, …,它是与下述 BL 模型等价的

$$X_t=\frac{1}{2}(1-\varepsilon_t+\varepsilon_t X_{t-1}),\qquad t=1,2,\cdots$$

其中 $\varepsilon_t=\pm1$的概率各为$\frac{1}{2}$, $X_0\sim U[0,1]$.

4. 检验非线性性

考虑最简单的一阶自归回模型:

$$X_t=f(X_{t-1})+\varepsilon_t,\qquad t=1,2,\cdots$$

其中 ε_t 满足: 给定序列过去历史之后的条件期望 $E(\varepsilon_t|X_{t-1},X_{t-2},\cdots)=0$. (这种 ε_t 还有一个名称叫鞅差平稳序列.)当 f 是非线性的这就是 NLAR 模型, 而当 $f(x)=c+\varphi_1 x$,c 为常数, φ_1 为回归系数时就是 AR(1) 模型. 问题是有了实际观察如何区

分?应该用什么模型?因而提出假设检验问题：

原假设：f 为线性的.

对立假设：f 为非线性的.

为避免技术细节，抓住方法的实质，我们宁肯多加一些条件以简化证明.

设 $\{X_t\}$ 为满足模型 $X_t = f(X_{t-1}) + \varepsilon_t$ 的平稳遍历的时间序列，而且 $EX_t = \mu$ 是已知的，自协方差函数 $r(\tau) = \mathrm{Cov}(X_t, X_{t-\tau}) = E(X_t - \mu)(X_{t-\tau} - \mu)$ 也是已知的，而且序列满足很好的矩存在条件及强混合条件. 记 $u_t = X_t - \mu$，(中心化)，$e_t = u_t - \rho u_{t-1}$，其中 $\rho = \rho(1) = r(1)/r(0)$ 为一步自相关函数. 显然有 $Eu_t = EX_t - \mu = 0$；$Ee_t = 0$ 以及 $Ee_t^2 = (1-\rho^2)r(0)$. 记 $Ee_t^2 = \sigma^2$，注意到 σ^2 为 $E(u_t - \alpha u_{t-1})^2$ 当 α 变化时的最小值（对 α 求导数即可证明当 $\alpha=\rho$ 时取到此极小值）. 故有

$$\sigma^2 = \inf_{\alpha} E(u_t - \alpha u_{t-1})^2.$$

引理 模型 $X_t = f(X_{t-1}) + \varepsilon_t$ 中，f 是线性函数的充分必要条件是 $E(e_t | u_{t-1}) = 0$ 对所有 $t \geqslant 1$ 几乎处处成立.

证明 充分性是显然的，仅证必要性.

若 $X_t = c + \phi_1 X_{t-1} + \varepsilon_t$，则 $X_t - \mu = c - (1-\phi_1)\mu + \phi_1(X_{t-1} - \mu) + \varepsilon_t$. 由 $Eu_t = 0$ 及 $Ee_t = E[E(e_t | X_s, s \leqslant t-1)] = 0$，知 $c = (1-\phi_1)\mu$，故 $u_t = \phi_1 u_{t-1} + \varepsilon_t$. 再由平稳性知 $|\phi_1| < 1$，又有

$$\phi_1 = \frac{Eu_{t-1}u_t}{Eu_{t-1}^2} + \frac{E\varepsilon_t u_{t-1}}{Eu_{t-1}^2} = \frac{Eu_t u_{t-1}}{Eu_{t-1}^2} = \rho.$$

这里利用了 $E(\varepsilon_t | X_s, s \leqslant t-1) = 0$，所以

$$e_t = u_t - \rho u_{t-1} = u_t - \phi_1 u_{t-1} = \varepsilon_t.$$

于是有（几乎处处）

$$E(e_t|u_{t-1}) = E(\varepsilon_t|u_{t-1})$$
$$= E[E(\varepsilon_t|X_s, s \leqslant t-1)|u_{t-1}] = 0. \quad \square$$

基于此引理我们可以引入 Kolmogrov-Smirnov 型的检验统计量

$$K_n = \sup_{-\infty<x<\infty} \left| \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} e_t 1_{(u_{t-1}<x)} \right|.$$

并有如下的

定理4.3.2 若 $X_t = f(X_{t-1}) + \varepsilon_t$ 中 f 为线性的，则必有 $\lim\limits_{n\to\infty} \frac{1}{\sqrt{n}} K_n$ 几乎处处大于0；又当原假设 f 是线性的成立时，有 $K_n \xrightarrow{\text{依分布}} K$，其中 $K = \sup\limits_{0\leqslant t\leqslant 1} |B(t)|$，$B(t)$ 为标准的[0,1]上的布朗桥运动.

证明 由引理知，当 f 非线性，$E(e_t|u_{t-1}) = 0$不能几乎处处成立. 故存在 x_0使 $Ee_2 1_{(u_1<x_0)} \neq 0$（否则由平移性可证引理条件成立），故亦有

$$Ee_t 1_{(u_{t-1}<x_0)} = Ee_2 1_{(u_1<x_0)} \neq 0.$$

由 $\{u_t\}$ 之遍历性，当 $n\to\infty$有（几乎处处）

$$\frac{1}{n}\sum_{k=2}^{n} e_t 1_{(u_{t-1}<x_0)} \longrightarrow Ee_2 1_{(u_1<x_0)}.$$

由 K_n 之定义有当 $n\to\infty$，

$$\frac{1}{\sqrt{n}} K_n \geqslant \left| \frac{1}{\sigma n} \sum_{k=2}^{n} e_t 1_{(u_{t-1}<x_0)} \right| \longrightarrow \frac{1}{\sigma} |Ee_2 1_{(u_1<x_0)}| > 0$$

几乎处处成立. 有关依分布收敛部分较复杂请参看安鸿志等人

的证明（刊国际统计评论1991年第3期）． □

上述结果的主要结论是在 f 为非线性时，当 $n\to\infty$ 时统计量 K_n 以 $\sqrt{n}$ 的速度趋于无穷．而当 f 为线性时，K_n 可以收敛到布朗桥最大值的分布．正是极限性状的不同使人们得以鉴别线性与非线性．换言之，当观察足够多时，经验的累积使人们能从偶然的表象中分析出精细的必然的差异．K 的分布在§4.1中已经给出．于是当有观察值 $\{X_1, \cdots, X_n\}$ 时，计算出 K_n．当 $K_n > K_\alpha$ 时，以水平 α 否定 f 的线性性．当 $\alpha=0.05$，$K_\alpha=2.24$；当 $\alpha=0.01$时，$K_\alpha=2.81$．这一方法也有多维的形式，是我国学者安鸿志、程斌完成的．其他检验非线性性的方法也还有许多，无法一一介绍．而这种利用大样本中收敛速度阶的区别来识别模型的方法在模型选取中也十分有用，例如 ZIC 准则．

5. 模型选择

非线性时间序列的模型选取也如同线性时间序列分析一样，可以借助于赤池的信息论准则，AIC 准则，其大意是若 f 为有待建模的真实的潜在结构．$g_m(\cdot|\theta)$，$m=1, \cdots, M$ 为用以建模的备选模型．模型带有未知参数 θ，参数个数为 r．基于观察值 x 作 θ 的极大似然估计（此估计将随 m 的不同而不同）．Akaike 提出在 $m=1, \cdots, M$ 中，取使

$$\log g_m(x|\hat{\theta}) - r$$

达到最大值的 m_0，以 $g_{m_0}(\cdot|\theta)$ 作为 f 的估计的最佳选择．前一项代表通常的对数似然，后一项起补偿作用．可以证明 AIC 准则实质上是要求与衡量两个分布间距离有关的一个量相对熵取最大值．f 与 g 的相对熵为

$$B(f;g)=-\int \frac{f(z)}{g(z)}\log \frac{f(z)}{g(z)}g(z)dz.$$

这是 Boltzman 热力学熵的一种推广. $B(f,g)$ 越大，$-B(f,g)$ 越小，g 与 f 则越接近. 特别当 $g=f$ 有 $-B(f,g)=0$. 如果取 $g_m(x|\hat{\theta})$ 为 $g(X_{t-1},X_{t-2},\cdots,X_{t-m},\varepsilon_t)$，则模型选取问题又化归为建模中极其重要的定阶问题. 而在混沌数据的时滞再造中嵌入维数的估计问题就是一个模型选取问题. NLAR 的一整套分析、辨识、定阶、诊断方法都可以为混沌现象的研究提供有力工具. 限于篇幅本书无法详细讨论分析的细节，读者可以求教于有关的专著.

6. 动力系统的稳定性与时间序列的遍历性

定义 一个动力系统称为是 Lagrange 稳定的，如果从初值 x_0 出发对所有 $k\geqslant 0$ f^k (x_0) 有界. 如果对所有初始值 $x\in R^m$ 均有 Lagrange 稳定，则称 f 是 Lagrange 稳定. 考虑满足如下差分关系的确定性模型

$$x_t=\begin{cases} ax_{t-1}, & \text{当} \quad x_{t-d}\leqslant 0, \\ bx_{t-1}, & \text{当} \quad x_{t-d}>0, \end{cases}$$

及随机的 TAR 模型

$$X_t=\begin{cases} a\,X_{t-1}+\varepsilon_{1,t}, & \text{当} \quad X_{t-d}\leqslant 0, \\ b\,X_{t-1}+\varepsilon_{2,t}, & \text{当} \quad X_{t-d}>0. \end{cases}$$

有趣的是 Chen，Tsay 1991年证明了上述 TAR 模型有几何遍历性的充要条件为 $a<1$，$b<1$，$ab<1$，$a^{o(d)}b^{e(d)}<1$，$a^{e(d)}b^{o(d)}<1$，其中 $e(d)$ 为依赖 d 的偶数，$o(d)$ 为依赖 d 的奇数. 之后又在1992年 Lim 证明了对确定性模型其 Lagrange 稳定性有完全一样的充要条件. 两类理论问题的内在联系有待研究.

确定性与随机模型都面临如何建立更好的符合实际的数学

模型的问题，它们之间也并没有截然分开的鸿沟．在有些系统中输入的是确定的信号，而输出却呈随机性．反过来一个随机性的输入却能产生（n 足够大！）完全确定的输出（回想羊齿草的例子），所以芝加哥大学的 Tsay 教授认为：混沌与传统统计学的重要区别在于后者发端于线性模型和 Gauss 分布，而前者从一开始就是非线性的．

§4.4 非参数回归

从一般模型 $Y_t=f\ (Y_{t-1},\ \eta_t)$ 的特例出发，当误差可加时就自然产生了非参数回归的问题．它将使人们能对潜结构 $f\ (\cdot)$ 作出估计，也能在此基础上估计映射 f 的 Lyapunov 指数，所以它将对混沌的研究发挥重要作用．下面作些粗浅介绍．

设 $\boldsymbol{X}$ 和 Y 分别为 d 维和一维的随机变量．

假定 Y 的期望 $E|Y|<\infty$，则 $f(x)=E(Y|\boldsymbol{X}=x)$ 存在，$f(x)$ 称为 Y 对 X 的回归函数．取 $Y=X_t,\boldsymbol{X}=(X_{t-1},\cdots,X_{t-d})$，则在模型 $X_t=f(X_{t-1},\cdots,X_{t-d})+\epsilon_t$ 中若 $E\epsilon_t=0,\epsilon_t$ 与 $X_s\ s<t$ 独立就有 $E(X_t|X_{t-1},\cdots,X_{t-d})=f(X_{t-1},\cdots,X_{t-d})$，正是我们感兴趣的映射 f（为方便计取消黑体 X）．回归分析的基本问题在于从（X，Y）抽取 i. i. d 样本（X_i，Y_i）（$i=1$，…，n）去估计回归函数 f．在传统的回归分析中，往往假定 $f(x)$ 有某种特定的数学形式，如线性型 $\beta_0+\sum_{i=1}^{d}\beta_iX_i$，并常常假定误差 $Y-f(x)$ 服从正态分布．此时经典的最小二乘法可以对 β_i 进行估计，并据此获得对 f 的估计．在模型假定条件满足的前提下，这些估计有许多优良的性质．但实际中 f 的线性性和误差的正态

性都未必能满足，于是只好另谋他途. 非参数回归就是重要选择之一. 这一模式足够广泛，如对估计条件概率（比如 Markov 过程中的转移函数），可定义 $Z=1_{[a,b]}(Y)$，则 $E(Z|X=x)$ 这一回归函数正是要估计的概率 $P\{a\leqslant Y\leqslant b|X=0\}$. 在非参数回归中 Y 与 X 都是随机的这也是它与通常回归模型的区别之点.

基于样本$(X_i,Y_i)(i=1,\cdots,n)$ 怎样来估计 $f(x)$ 呢?权函数是一个十分有用的方法. 它的大意是把样本 (X_i,Y_i) 中使 X_i 恰好等于指定 x 的那些样本 $X_{i_1},\cdots,X_{i_k}$挑出来，以其自然对应的 Y_{i_1}，…，Y_{i_k}的算术平均值 $\frac{1}{k}\sum_{j=1}^{k}X_{i_j}$ 作为 f 在 x 点的值的估计. 但一般恰 好为 x 的 X_i 未必有那么多，于是找一个充分小 $h_n>0$，考虑 X_1，…，X_n 中落在区间 $[x-h_n,\ x+h_n]$ 内的，比方说仍为 $X_{i_1},\cdots,X_{i_k}$，则仍以上述算术平均去估计 $f(x)$. 因此这个估计一般可写为 $\sum_{i=1}^{n}W_{ni}Y_i$，其中

$$W_{ni}=\begin{cases}\frac{1}{k}, & \text{当 } i \text{ 为 } i_j,j=1,\cdots,k \text{ 之一},\\ 0, & \text{其他 } i.\end{cases}$$

注意到权 W_{ni}与 x 及 X_1，…，X_n 有关. 因此也可称为 X_i 的权. 更确切地说，是在整个样本中相对于 x 点的权，它反映了在估计 $f(x)$ 时样本(X_i,Y_i) 的作用有多大.

上述权只是一个特例，

一般可以

定义 设 $W_{ni}=W_{ni}(x)=W_{ni}(x;X_1,\cdots,X_n)$，$i=1,2,\cdots,n$，是选定的 n 个依赖于 x 和 X_1，…，X_n 的函数，则

$$f_n(x) = \sum_{i=1}^{n} W_{ni} Y_i$$

为 $f(x)$ 的权函数估计，$\{W_{ni}\}$ 称为权函数. 作为权，自然要求

$$W_{ni} \geqslant 0, \quad \sum_{i=1}^{n} W_{ni} = 1.$$

权函数方法估计 f 的大样本理论（或者说极限理论）由 Stone 在 1977年作了奠基工作，之后研究十分活跃. 从实用观点看，重要的是如何选择适当的权函数. 常用的权函数选取法为

a）核函数方法.

取 R^d 上的函数 K，一般取为概率密度及窗宽 $h_n>0$，然后令

$$W_{ni} = K\left(\frac{x - X_i}{h_n}\right) \Big/ \sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right).$$

这种权最早为 Watson 等所考虑. 这与一般概率密度估计中的权函数选取十分类似.

b）近邻方法.

该法先引入 R^d 中的距离$\|\cdot\|$（距离的引入也颇有讲究，无法细述）. 设有了样本 (X_i, Y_i)，$i=1, \cdots, n$，指定 R^d 中的一点 x，将 $X_1, \cdots, X_n$ 按在$\|\cdot\|$定义下与 x 接近的程度排序：

$$\|X_{R_1} - x\| < \|X_{R_2} - x\| < \cdots < \|X_{R_n} - x\|,$$

再引入一组 C_{ni} $i=1, \cdots, n$ 满足 $C_{n1} \geqslant C_{n2} \geqslant \cdots \geqslant C_{nn}$，$\sum_{i=1}^{n} C_{ni} = 1$，对离 x 最近的 X_{R_1} 赋以权 C_{n1}，依次类推，最后有

$$W_{nR_i}(x; X_1, \cdots, X_n) = C_{ni}, \quad i = 1, \cdots, n.$$

当有点与 x 距离相等时对权在他们之间平均分配. 理论已证明近邻权是一个有优良性质的权. 在此基础上，继 Devroye (1981) 等人之后，陈希孺，赵林城等在相当广泛的条件下对核

函数权和近邻权都证明了当 $n\to\infty$ 时，几乎处处有

$$\lim_{n\to\infty} f_n(X) = f(X).$$

详细结果可参看前面提及的陈希孺的著作. 由此可见统计学家可借助非参数回归为寻找迭代机制作贡献.

在混沌学的研究中还有几点使之必须面对统计问题：首先初始值往往是未知的，也有人认为其随机行为的来源正是初始值的不易确定，所以需对初始值的分布作研究. 而在实际物理过程的研究中，观察、计算工具都不可避免地有误差. 因而真实世界的模型与其说是 $X_t = f(X_{t-1})$ ，不如说是 $X_t = f(X_{t-1}, \eta_t)$. 另一方面人们对过程的观察在时间与空间两方面都受到限制，类似时滞再造的方法至关重要，这些问题都是统计学家所必须面对的.

学科间的影响是相互的，混沌科学的蓬勃发展也对其他学科包括数学的各分支，当然也包括统计带来各种影响：重要的有关于非线性问题，非线性统计方法的发展；长周期，长记忆模型的研究；分数差分的 ARIMA 模型及分数布朗运动都是分形思想在统计学中的印迹. 学科的渗透将促进学科的发展. 借助于物理学、生物学、地球科学、气象地震科学、计算机科学、高分子化学、材料科学、当然还有数学乃至统计数学，人类将认识混沌，走出混沌.

附　记

对想以统计方法作为工具来认识混沌的读者．以及对此有兴趣的统计研究者以下文献可使你了解最近的研究进展．

1. 1989年 Edinburgh 国际非线性时间序列会议的一组论文刊登在泛华统计杂志 Statistica Sinica Vol. 1 和2（1991，1992）．
2. 1990年英国皇家统计学会会议讨论 Bartlett 的报告“机会或混沌？”的一组文章刊登在 J. R. Statist. Soc. A（1990）．
3. 1991年“英国皇家统计学会混沌日”会上的一组论文，刊登在 J. R. Statist. Soc. B Vol. 54（1992）．
4. 1992年美国“统计科学”杂志“两篇关于统计与混沌理论论文”的笔谈讨论，刊于 Statistical Science Vol. 7.（1992）．

各组论文的后面像随机瀑布一样有丰富的精细结构和读不尽的前人成果．

写完这本小册子，回想其主要内容只是以遍历性定理为主线讲述了统计学者对随机性的看法，关于混沌科学则仅涉及一点皮毛，算是一种读书札记，离“走向数学”丛书的宗旨尚有很大距离．但有一点我是确信的：走向数学才能走出混沌．

作者感谢冯克勤教授的支持，同时作者还借此机会感谢参考文献的作者们．

作者

1994，5　识于合肥

参 考 书 目

1. Peitgen, Jürgens, Saupe, Chaos and Fractals. New Frontiers of Science. Springer-Verlag. 1992.
2. Howell Tong, Non-linear Time Series A Dynamical System Approach. Clarendon Press Oxford 1990.
3. Howell Tong ed. Dimension Estimation and Models, World Scientific. 1993.